

Le codage du flot géodésique sur la surface modulaire

Pierre ARNOUX^{1,2}

Résumé. – Nous donnons une preuve élémentaire et explicite du fait que le flot géodésique sur la surface modulaire (quotient du plan hyperbolique par l'action de $SL(2, \mathbb{Z})$) peut être codé en utilisant les fractions continues. (Réédition avec autorisation de l'auteur et de la revue de l'article « P. Arnoux, Le codage du flot géodésique sur la surface modulaire, *Enseign. Math.* (2) 40, no. 1-2, 29–48, 1994 ».)

Abstract. – We give an elementary and explicit proof of the coding of the geodesic flow on modular surface by continued fractions.

1. Introduction

Il est connu depuis longtemps, par un travail d'Artin (cf. [Art24]), que le flot géodésique sur la surface modulaire peut être codé en utilisant les fractions continues; les articles récents d'Adler et Flatto et de Series ont fait une étude approfondie de ce codage (cf. [AF91], [Ser85]). Le but de cet article est de retrouver ce résultat de façon explicite et élémentaire en mettant un système de coordonnées adapté sur le fibré unitaire tangent de la surface modulaire.

De façon plus précise, on peut définir algébriquement le flot géodésique sur la surface modulaire comme l'action à droite, sur l'espace quotient $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$, du groupe des matrices diagonales positives de la forme

$$g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$$

(nous rappellerons plus bas la démonstration); mais on peut aussi voir $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ comme l'espace des réseaux de \mathbb{R}^2 dont le domaine fondamental est de volume 1, et l'action de g_t consiste alors à écraser le réseau le long de l'axe des abscisses, en multipliant les abscisses par $e^{\frac{t}{2}}$ et les ordonnées par $e^{-t/2}$.

Pour la plupart des réseaux (ceux qui ne contiennent pas de vecteurs horizontaux ou verticaux), on peut trouver un domaine fondamental en forme de L, formé de deux rectangles accolés; l'action du flot dilate les bases de ces domaines fondamentaux et écrase leurs hauteurs, on peut alors revenir à un domaine de base plus petite en découpant le grand rectangle et en empilant les morceaux obtenus au-dessus du petit rectangle (cf. Fig. 1). Si l'on part de deux rectangles de bases respectives a et $b < a$, l'opération initiale est la division euclidienne de a par b , le quotient apparaît dans le nombre d'empilements au-dessus de b , puis on reprend l'opération avec b et le reste de la division, de manière identique à l'algorithme d'Euclide; mais ici l'algorithme ne termine jamais dès que les

1. GDAC-I2M UMR 7373 CNRS Université d'Aix-Marseille. Au moment de la publication (d'après le manuscrit du 23 août 1993) : Laboratoire de Mathématiques Discrètes (UPR 9016), 163 avenue de Luminy, Case 930, 13288 Marseille Cedex 9, France)

2. pierre.arnoux@univ-amu.fr

deux nombres sont incommensurables : il s'agit en fait de la version vectorielle de l'algorithme classique des fractions continues, et un calcul simple montre que si les quotients successifs obtenus sont a_1, a_2, \dots , on a :

$$\frac{b}{a} = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\ddots}}}}$$

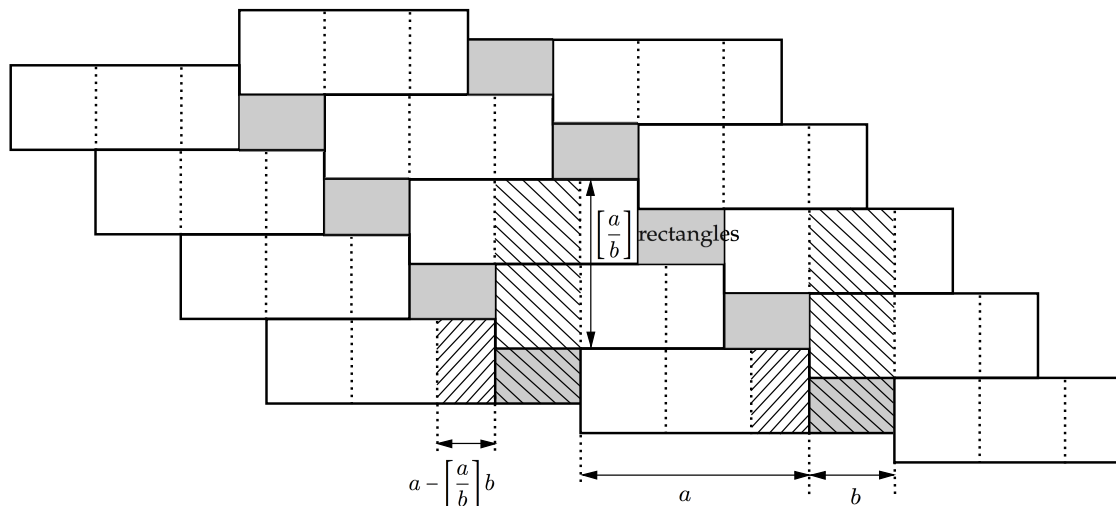


FIGURE 1

Il est clair que rien n'est changé si l'on multiplie a et b par la même constante, on peut donc à chaque étape normaliser à 1 le plus grand des deux nombres, et on passe ainsi de b à la partie fractionnaire de $1/b$, ce qui est la fonction associée à l'algorithme des fractions continues. Puisque le flot g_t dilate les longueurs, on peut utiliser un temps convenable du flot pour réaliser cette normalisation. La suite de cet article est consacrée à rendre rigoureux le raisonnement qui précède, et à en tirer quelques conséquences. Une façon beaucoup plus lourde d'énoncer les résultats qui précèdent consiste à voir l'espace $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ comme espace modulaire du tore, et le flot g_t comme flot de Teichmüller sur cet espace; ce point de vue, inutile ici, permet d'étudier le flot de Teichmüller sur l'espace modulaire d'une surface quelconque, en codant par des « rectangles cousus », généralisations de notre domaine fondamental en L. Cette étude a été faite par Veech (cf. [Vee86]), et l'on peut considérer le présent article comme un exposé du cas le plus simple de cette construction (cf. [Vee84, p. 1391], où on étudie aussi le flot horocyclique).

Dans la première section, nous rappelons ce que nous aurons besoin de savoir sur les fractions continues; dans la deuxième section, nous définissons le flot géodésique sur la surface modulaire, et nous en donnons un autre modèle dans la troisième section. Dans la quatrième, nous définissons de manière élémentaire un système de coordonnées global sur le fibré unitaire tangent à la surface modulaire, et dans la cinquième section, nous montrons que le flot géodésique admet une section sur laquelle l'application de premier

retour est un revêtement double de l'extension naturelle de la transformation des fractions continues. Dans la sixième section, nous donnons, comme exemple d'application de cette méthode, une démonstration géométrique d'un théorème de Paul Lévy sur la croissance des dénominateurs des convergents pour presque tout nombre, en utilisant l'ergodicité du flot géodésique. Dans la dernière section, nous montrons comment, en changeant la section choisie pour le flot géodésique, on peut retrouver l'algorithme additif des fractions continues; on pourrait ainsi trouver, avec d'autres sections, une infinité d'algorithmes du même type.

2. La transformation des fractions continues

Il est classique que tout réel $x \in]0, 1[$ s'écrit de façon unique

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\ddots}}}}$$

où la suite (a_n) est une suite d'entiers strictement positifs, finie et se terminant par un entier strictement plus grand que 1 si x est rationnel, infinie sinon. On note habituellement cette égalité $x = [0; a_1, \dots, a_n, \dots]$; dans le cas particulier où la suite est périodique, on note $x = [0; \overline{a_1, \dots, a_n}]$. Un calcul immédiat montre que $a_1 = [1/x]$, où $[t]$ est la partie entière de t , et que, si l'on pose $x_1 = \{1/x\}$, où $\{t\} = t - [t]$ est la partie fractionnaire de t , on a $a_2 = [1/x_1]$. Plus généralement, définissons les deux applications

$$\begin{aligned} f :]0, 1[&\rightarrow \mathbb{N}, & x &\mapsto [1/x] \\ T : [0, 1[&\rightarrow [0, 1[, & x &\mapsto \{1/x\} \text{ si } x \neq 0, \quad T(0) = 0 \end{aligned}$$

On vérifie que a_n est défini si $T^{n-1}(x) \neq 0$, et vaut alors $f(T^{n-1}(x))$.

La transformation T est appelée transformation des fractions continues, elle est surjective, mais non injective; on peut dans un tel cas construire une application bijective associée, "l'extension naturelle", par limite inductive, en considérant tous les passés possibles pour un point, c'est-à-dire en considérant les suites $(x_i)_{i \in \mathbb{Z}}$ telles que $T(x_i) = x_{i+1}$ (voir [Roh64] pour une étude plus poussée des extensions naturelles).

Cette extension naturelle admet ici, à un ensemble de mesure nulle près, un modèle géométrique : une application \bar{T} définie sur un sous-ensemble de \mathbb{R}^2 , bijective sauf sur un ensemble de mesure nulle, et telle que, si l'on note π la projection sur la première variable, on ait $\pi(\bar{T}(x, y)) = T(x)$, la deuxième variable représentant en quelque sorte le "passé". Cette application a pour ensemble de définition le sous-ensemble Σ du plan défini par les équations $0 \leq x < 1$, et $0 \leq y < 1/(1+x)$ si $x < 1/2$, $0 < y < 1/(1+x)$ sinon, et elle est définie par

$$\begin{aligned} \bar{T} : \Sigma &\rightarrow \Sigma, & (x, y) &\mapsto (\{1/x\}, x - x^2y) \text{ si } x \neq 0 \\ & & (0, y) &\mapsto (0, 0) \end{aligned}$$

Cette application est bijective sauf aux points $(0, y)$ qui ont tous même image et aux points $(x, 0)$ qui n'ont pas d'antécédent si $x \neq 0$; elle envoie les segments verticaux sur des segments verticaux, et elle est discontinue sur les segments $x = 1/n$ pour n entier, mais continue sur les rectangles curvilignes qu'ils délimitent. Un calcul simple montre que le rectangle curviligne défini par $1/(n+1) < x < 1/n$, $0 < y < 1/(1+x)$ est envoyé sur

le rectangle défini par $0 < x < 1$, $1/(n+1+x) < y < 1/(n+x)$, avec une rotation d'un demi-tour, qui envoie le bord droit sur le bord gauche et le bord inférieur sur le bord supérieur. C'est pour cela qu'il faut enlever de Σ les points $(x, 0)$, $x \geq 1/2$, qui seraient envoyés sur le bord supérieur qui n'appartient pas à Σ (voir Fig. 2).

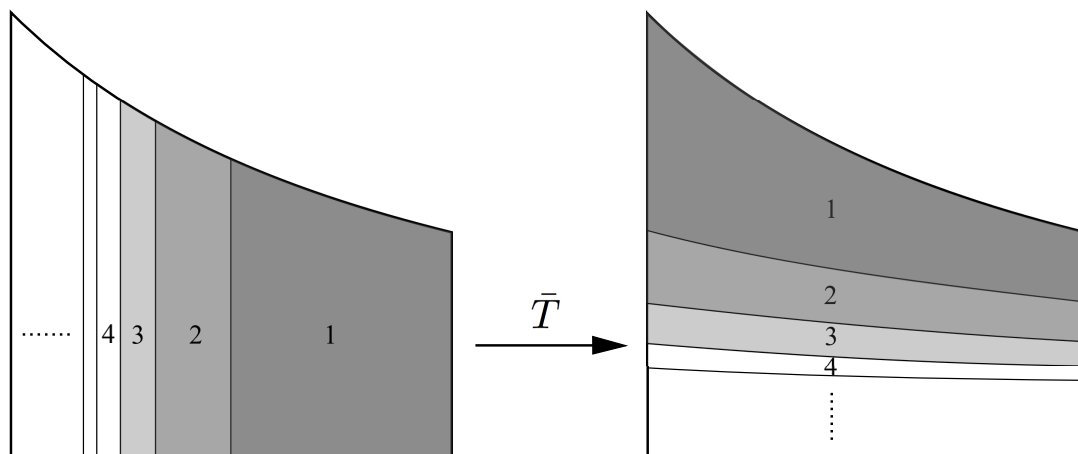


FIGURE 2

C'est l'application \bar{T} que nous trouverons naturellement ci-dessous comme application de premier retour du flot géodésique sur la surface modulaire; elle a de plus l'intérêt de préserver la mesure de Lebesgue, puisque son jacobien est 1 en tout point où elle est continue; la projection sur la première coordonnée donne la mesure de Gauss, de densité $1/(1+x)$, préservée par T (pour plus de détails, voir [AN]).

On peut présenter cette application d'une autre manière : le changement de variable $(x, y) \mapsto (x, y/(1-xy))$ envoie Σ sur le carré $[0, 1] \times [0, 1]$ privé d'une partie de son bord; on vérifie facilement qu'il conjugue l'application \bar{T} à l'application \bar{T}_1 définie, sur l'intérieur de son domaine, par

$$\bar{T}_1 :]0, 1[\times]0, 1[\rightarrow]0, 1[\times]0, 1[, \quad (x, y) \mapsto \left(\left\{ \frac{1}{x} \right\}, \frac{1}{y + \left\lfloor \frac{1}{x} \right\rfloor} \right)$$

C'est sous cette forme que l'on trouve en général l'extension naturelle de la transformation des fractions continues (voir par exemple [Nak81], [Ito84]). Cette application est continue sur les rectangles d'équation $1/(n+1) < x < 1/n$, qui sont envoyés respectivement sur les rectangles $1/(n+1) < y < 1/n$; ces rectangles forment une partition de Markov, et il est particulièrement commode d'utiliser \bar{T}_1 pour construire un **système dynamique symbolique**, de la façon suivante : A chaque couple (x, y) de nombres irrationnels de l'intervalle $[0, 1]$, on associe la suite $(a_n)_{n \in \mathbb{Z}}$ définie par $x = [0; a_1, \dots, a_n, \dots]$ et $y = [0; a_0, a_{-1}, \dots, a_{-n}, \dots]$; Ceci définit une bijection entre l'ensemble I des points du carré de coordonnées irrationnelles et l'ensemble des suites biinfinies à valeurs entières strictement positives. On vérifie immédiatement que I est le complémentaire, dans le carré $]0, 1[\times]0, 1[$, des orbites (pour \bar{T}_1) des points de la forme $(x, 0)$ ou $(0, y)$. Cet ensemble est de mesure pleine, invariant par \bar{T}_1 , et la restriction de \bar{T}_1 à I est une bijection.

De même que l'on a, pour le développement en fraction continue, $a_n = f(T^{n-1}(x))$ avec n entier positif, on a ici $a_n = \bar{f}(\bar{T}_1^{n-1}(x, y))$ pour tout n dans \mathbb{Z} , où l'on définit \bar{f} par $\bar{f}(x, y) = [1/x]$ (numéro du rectangle qui contient le point (x, y)). Cette formule montre que l'application de codage qui à (x, y) associe la suite $(a_n)_{n \in \mathbb{Z}}$ conjugue \bar{T}_1 au décalage S sur $\mathbb{N}^*\mathbb{Z}$ (S est l'application qui à la suite $(u_n)_{n \in \mathbb{Z}}$ associe la suite $(v_n)_{n \in \mathbb{Z}}$ définie par $v_n = u_{n+1}$).

3. Le flot géodésique sur la surface modulaire

Du point de vue algébrique, le flot géodésique sur la surface modulaire est l'action à droite sur le quotient $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ du groupe des matrices diagonales positives, de forme

$$g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$$

Pour justifier ce nom de flot géodésique, il faut rappeler un peu de géométrie. On nomme demi-plan de Poincaré l'ensemble $\mathbb{H} = \{x + iy \mid y > 0\}$ des complexes à partie imaginaire strictement positive, muni de la métrique $(dx^2 + dy^2)/y^2$. Le groupe $SL(2, \mathbb{R})$ agit à gauche sur \mathbb{H} par

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \cdot z = \frac{az + c}{bz + d}$$

et cette action est isométrique. On vérifie facilement qu'elle est transitive, et que le stabilisateur du point i est le groupe $SO(2, \mathbb{R})$, donc le demi-plan de Poincaré s'identifie à l'espace homogène $SL(2, \mathbb{R})/SO(2, \mathbb{R})$. Le stabilisateur agit transitivement sur le cercle unité du plan tangent à i , donc $SL(2, \mathbb{R})$ agit transitivement sur le fibré unitaire tangent du demi-plan. Le stabilisateur du vecteur tangent vertical en i est $\{Id, -Id\}$ (l'action de la matrice $-Id$ sur \mathbb{H} est triviale), et le fibré unitaire tangent s'identifie au quotient $PSL(2, \mathbb{R})$ de $SL(2, \mathbb{R})$ par son centre $\{\pm Id\}$. Par raison de symétrie, la courbe

$$\gamma(t) = e^t i = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \cdot i$$

est une géodésique, et un calcul facile montre qu'elle est paramétrée par sa longueur (c'est de là que vient le facteur $1/2$ dans la définition de g_t). Puisqu'une isométrie transforme géodésique en géodésique, la courbe $M \cdot \gamma(t)$ est aussi une géodésique, on en déduit que le flot géodésique du demi-plan de Poincaré se représente matriciellement par l'action à droite sur $PSL(2, \mathbb{R})$ du groupe $\{g_t \mid t \in \mathbb{R}\}$.

On vient de voir que les géodésiques de \mathbb{H} sont les images d'une droite verticale par les isométries; le groupe des isométries est engendré par des translations horizontales et des inversions centrées sur l'axe des abscisses, ce qui permet de retrouver un fait classique : les géodésiques du demi-plan de Poincaré sont les demi-droites verticales et les demi-cercles centrés sur l'axe.

La surface modulaire est le quotient du demi-plan par l'action du groupe $PSL(2, \mathbb{Z})$, qui est engendré, en tant que groupe agissant sur \mathbb{H} , par les applications $z \mapsto z+1$, de domaine fondamental $-1/2 \leq z < 1/2$, et $z \mapsto -1/z$, de domaine fondamental (au bord près) $|z| \geq 1$;

on en déduit qu'un domaine fondamental pour l'action du groupe est contenu dans l'intersection de ces deux domaines, et l'on peut montrer qu'au bord près, cette intersection est en fait un domaine fondamental (cf. Fig. 3).

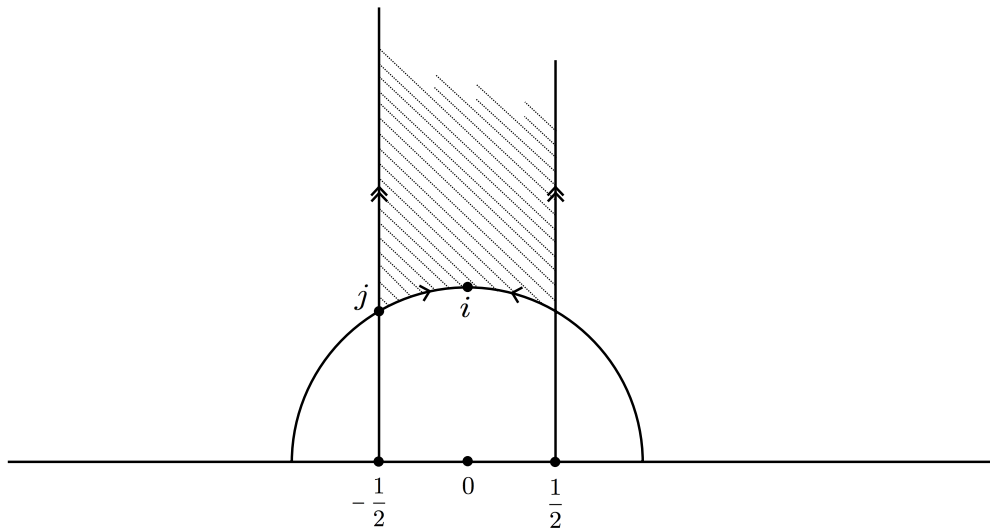


FIGURE 3. Le domaine fondamental pour l'action de $SL(2, \mathbb{Z})$.

Cette surface est isomorphe à une sphère privée d'un point avec 2 singularités. Son fibré unitaire tangent est isomorphe à $PSL(2, \mathbb{Z}) \backslash PSL(2, \mathbb{R}) \cong SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$, et on retrouve bien pour le flot géodésique la forme annoncée au début de la section.

4. Une autre présentation du flot géodésique

Nous allons donner une autre interprétation de ce flot : on peut considérer une matrice

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

de $SL(2, \mathbb{R})$ comme formée de deux vecteurs lignes $e_1^t = (a, c)$ et $e_2^t = (b, d)$. On lui associe ainsi une base (e_1, e_2) de \mathbb{R}^2 orientée dans le sens positif et dont le « carré unité » $\{xe_1 + ye_2, 0 \leq x < 1, 0 \leq y < 1\}$ est de volume 1. La multiplication à droite de cette matrice par une matrice M correspond, si l'on considère les vecteurs, à l'action standard sur le plan \mathbb{R}^2 de la matrice transposée M^t . Le flot géodésique consiste à « aplatiser » la base donnée sur l'axe des abscisses, et l'espace quotient $SL(2, \mathbb{R})/SO(2, \mathbb{R})$ s'interprète comme espace des bases à isométrie près. On peut facilement retrouver le modèle de Poincaré : si nous assimilons \mathbb{R}^2 à \mathbb{C} , on peut associer à une base (e_1, e_2) un couple (z_1, z_2) de nombres complexes, et définir une application

$$\begin{aligned} \phi & : SL(2, \mathbb{R}) \rightarrow \mathbb{H} \\ (e_1, e_2) & \mapsto z = z_2/z_1. \end{aligned}$$

Il est clair que ϕ est invariante par isométrie, et son image appartient au demi-plan supérieur ; c'est une version explicite de l'identification donnée au paragraphe précédent entre le demi-plan de Poincaré et le quotient $SL(2, \mathbb{R})/SO(2, \mathbb{R})$.

La multiplication à gauche par un élément

$$\begin{pmatrix} a & \beta \\ \gamma & \delta \end{pmatrix}$$

de $SL(2, \mathbb{Z})$ remplace (e_1, e_2) par $(ae_1 + \beta e_2, \gamma e_1 + \delta e_2)$, qui est une autre base du même réseau $\mathbb{Z}e_1 + \mathbb{Z}e_2$, et l'on obtient de cette façon toutes les bases directes de ce réseau. Le quotient à gauche $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ est donc l'espace des réseaux.

Pour pouvoir décrire explicitement le flot géodésique, il nous faut disposer de coordonnées sur cet espace des réseaux, c'est-à-dire donner une façon de choisir une base pour un réseau donné. La méthode classique est celle de Minkowski, qui consiste à choisir pour e_1 le plus court vecteur du réseau (bien défini au signe près, sauf si le réseau est carré ou hexagonal), et pour e_2 le plus court vecteur du réseau qui soit indépendant de e_1 et forme une base positivement orientée (bien défini, sauf si le réseau est hexagonal). Cette méthode est évidemment équivariante par toute isométrie, et il est facile de vérifier que les bases obtenues se projettent, via l'application ϕ , sur le domaine fondamental de l'action de $SL(2, \mathbb{Z})$ sur \mathbb{H} que nous avons défini dans la section 2. L'action triviale de $-Id$ sur le fibré unitaire tangent correspond au fait que la méthode de Minkowski donne pour chaque réseau deux bases opposées l'une de l'autre ; les réseaux pour lesquels cette méthode fournit plus de deux bases (réseaux carrés et hexagonaux) sont projetés sur les deux points singuliers de la surface modulaire.

5. Un système de coordonnées sur l'espace des réseaux

Nous allons définir une autre méthode pour fixer des coordonnées, qui conduit à des calculs plus faciles. Etant donné un réseau dans \mathbb{R}^2 (identifié à \mathbb{C} pour la commodité de notation), on définit pour chaque point x du réseau l'intervalle horizontal ouvert $H_x =]x - 1, x + 1[$. La demi-droite verticale issue de x recoupe au moins un autre intervalle horizontal : en effet, soit le réseau contient un vecteur vertical, et cette demi-droite rencontre un autre point du réseau, soit il n'en contient pas, donc les deux vecteurs d'une base du réseau ont des abscisses rationnellement indépendantes, et il y a des points du réseau arbitrairement proches de la demi-droite.

Soit H_y le premier intervalle rencontré, et supposons que le point de rencontre soit à gauche de y . Ce point de rencontre peut s'écrire $x + ic = y - a$; on appellera V_x l'intervalle vertical $[x, x + ic[$ qu'il définit. On prolonge ensuite vers la droite tous les intervalles horizontaux jusqu'à rencontrer un intervalle vertical, ce qui est toujours possible par un raisonnement du même type que ci-dessus. Par symétrie, pour tout intervalle V_x , il existe un seul H_z dont le prolongement le coupe; on écrira $z + b = x + id$ le point d'intersection. Il est clair que (a, c) et $(-b, d)$ sont des vecteurs du réseau, et on vérifie sans peine qu'ils forment une base (cf. Fig. 4; en fait, cette construction donne un domaine fondamental pour le réseau formé de deux rectangles, l'un de base a et de hauteur d , l'autre de base b et de hauteur c). On vérifie immédiatement que l'on a, par construction, $0 < a < 1 \leq b$ et $0 \leq d < c$.

Si le premier point de rencontre avec H_y est à droite de y , on peut faire la construction symétrique, et on obtient une base $(a, c), (-b, d)$ avec $0 < b < 1 \leq a$ et $0 \leq c < d$.

Cette construction tombe en défaut dans deux cas :

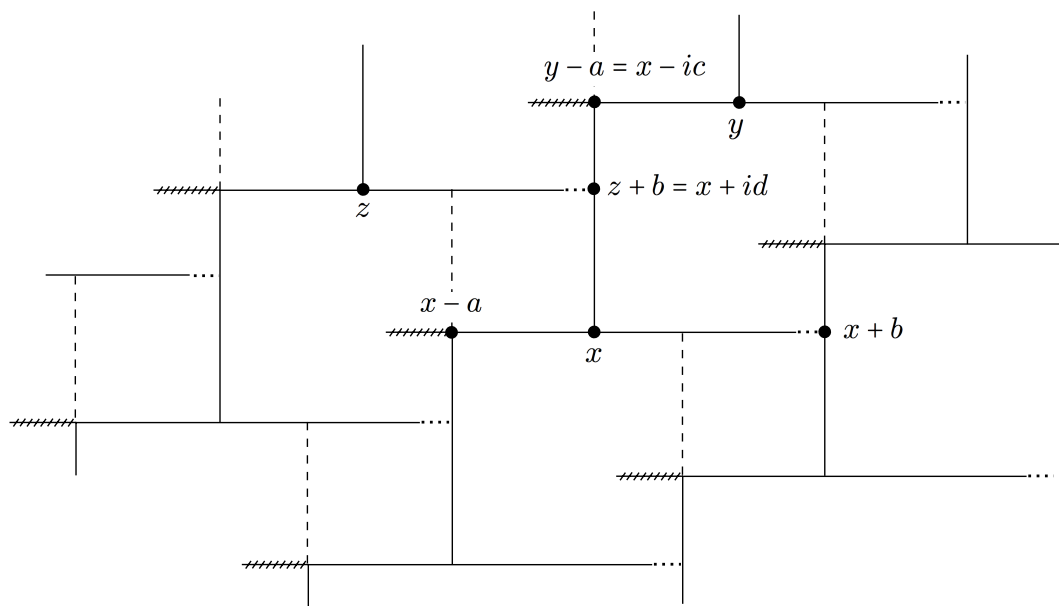


FIGURE 4.

- si le premier point d'intersection est égal à y ; dans ce cas le réseau contient un vecteur vertical, et on peut trouver une base de la forme $(0, c), (-b, d)$ avec $d < c, b \geq 1$ (et aussi une base de la forme $(a, c), (0, d)$ avec $c < d, a \geq 1$);
- si le premier point d'intersection est contenu dans deux intervalles horizontaux, et dans ce cas le réseau contient un vecteur horizontal, et il y a une base de la forme $(a, 0), (-b, d)$, avec $b < a < 1$.

On peut remarquer que le premier cas correspond à des points dont l'orbite positive tend vers l'infini sur la surface modulaire, et le second cas à des points dont l'orbite négative tend vers l'infini; par exemple, une base de la forme $(0, c), (-b, d)$ devient $(0, ce^{-t/2}), (-be^{t/2}, de^{-t/2})$ sous l'action du flot géodésique, et l'application ϕ définie ci-dessus lui fait correspondre le nombre complexe $d/c + ie^{t/2}b/c$, qui est contenu dans le domaine fondamental donné ci-dessus dès que $d/c < 1/2$ et tend vers l'infini.

Pour résumer, nous avons montré qu'il existe un domaine fondamental pour l'action à gauche de $SL(2, \mathbb{Z})$ sur $SL(2, \mathbb{R})$ formé de matrices $\begin{pmatrix} a & c \\ -b & d \end{pmatrix}$, où (a, b, c, d) appartient à l'un des 3 ensembles suivants :

$$\begin{aligned} \Omega_0 &= \{(a, b, c, d) \in \mathbb{R}^4 \mid ad + bc = 1, 0 < b < 1 \leq a, 0 \leq c < d\} \\ \Omega_1 &= \{(a, b, c, d) \in \mathbb{R}^4 \mid ad + bc = 1, 0 \leq a < 1 \leq b, 0 \leq d < c\} \\ \Omega_2 &= \{(a, b, 0, d) \in \mathbb{R}^4 \mid ad = 1, 0 < b < a < 1\}. \end{aligned}$$

Le choix pour les inégalités strictes ou larges sur les bords de Ω_0 et Ω_1 est assez arbitraire, puisqu'il y a des identifications entre les bords; en particulier le bord $a = 0$ de Ω_0 s'identifie au bord $b = 0$ de Ω_1 , l'identification étant donnée par la formule :

$$\begin{pmatrix} 0 & \frac{1}{a} \\ -a & \frac{1}{a} - c \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a & c \\ 0 & \frac{1}{a} \end{pmatrix}$$

Il peut être utile d'interpréter ces domaines de la façon suivante : tout réseau ne contenant pas de vecteur horizontal possède un unique domaine fondamental formé de deux rectangles alignés, le plus étroit étant de largeur inférieure à 1, le plus large étant plus haut que l'autre, et de largeur supérieure à 1 (cf. Fig. 4). On peut alors prendre comme coordonnées les deux largeurs et la plus petite des hauteurs ; l'autre s'en déduit puisque l'aire du domaine fondamental est 1. On retrouve le fait, évident sur les équations, que les domaines Ω_0 et Ω_1 sont de dimension 3 ; Ω_0 (resp. Ω_1) correspond au cas où c'est le rectangle de droite (resp. de gauche) qui est le plus grand. Quant à Ω_2 , il correspond aux réseaux contenant un vecteur horizontal petit ; le plus petit rectangle est alors de hauteur nulle, et l'on ne peut plus assurer que la largeur du plus grand rectangle soit supérieure à 1.

6. Le codage du flot géodésique

Dans ces coordonnées, le flot géodésique s'écrit simplement, du moins localement :

$$g_t(a, b, c, d) = (e^{t/2}a, e^{t/2}b, e^{-t/2}c, e^{-t/2}d)$$

Mais pour t assez grand, les deux premières coordonnées sont plus grandes que 1 (sauf dans le cas particulier où l'une d'entre elles est initialement nulle), et l'on traverse le bord du domaine fondamental ; il faut alors faire une identification pour poursuivre l'orbite à partir d'une autre face. Pour décrire complètement le flot, il faut étudier cette identification.

Le domaine Ω_0 peut être paramétré par les 3 coordonnées a, b, c , puisque $d = (1 - bc)/a$, et plongé dans \mathbb{R}^3 (cf. Fig. 5) ; il possède alors cinq bords, dont trois, donnés respectivement par les équations $b = 0, c = 0, c = 1/(a + b)$, sont formés de segments d'orbites du flot géodésique. Les deux derniers sont Σ_0 , d'équation $a = 1$, sur lequel le flot est rentrant, et Λ_0 , d'équation $b = 1$, sur lequel le flot est sortant. Le bord Σ_0 peut être paramétré par (b, c) , avec $b \in [0, 1[$ et $c \in [0, 1/(b + 1)[$; on reconnaît le domaine Σ du paragraphe 1. Le bord Λ_0 peut, lui, être paramétré par (a, c) , avec $a > 1$ et $c \in [0, 1/(a + 1)[$.

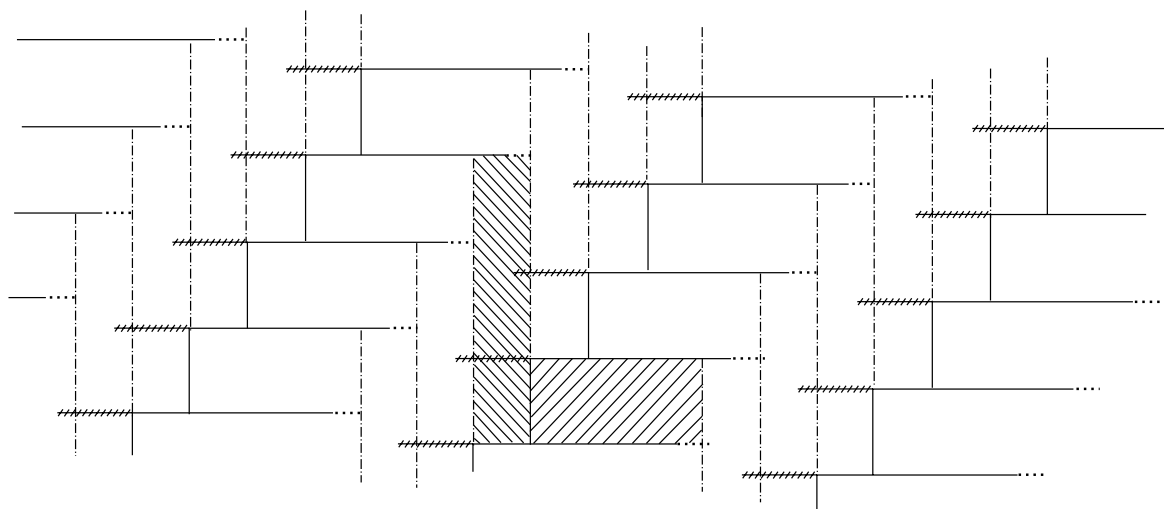


FIGURE 5.

On peut de même paramétrer Ω_1 , cette fois par (a, b, d) , et définir un bord rentrant Σ_1 , paramétré par (a, d) , et un bord sortant Λ_1 , paramétré par (b, d) .

Notons $\tilde{\Sigma} = \Sigma_0 \cup \Sigma_1$ la réunion des deux bords rentrant; on peut identifier $\tilde{\Sigma}$ à $\Sigma \times \{0, 1\}$, avec coordonnées (x, y, ϵ) où ϵ appartient à $\{0, 1\}$, (x, y, ϵ) correspondant à l'élément (x, y) de Σ_ϵ . Le résultat principal de cet article est le suivant :

Proposition 6.1. *L'application de premier retour \tilde{T} du flot géodésique en $\tilde{\Sigma}$ est un revêtement d'ordre 2 de l'application \bar{T} définie au paragraphe 1, c'est-à-dire qu'elle est donnée par*

$$\tilde{T} : \tilde{\Sigma} = \Sigma \times \{0, 1\} \rightarrow \tilde{\Sigma}, \quad (x, y, \epsilon) \mapsto (\{1/x\}, x - x^2y, 1 - \epsilon)$$

Preuve. La remarque essentielle est que Λ_0 s'identifie naturellement à Σ_1 ; en effet, le point (a, c) de Λ_0 s'identifie au point $(\{a\}, \frac{1-c}{a})$ de Σ_1 , par produit par un élément de $SL(2, \mathbb{Z})$, en tenant compte du fait que dans la formule suivante on a $d = \frac{1-c}{a}$:

$$\begin{pmatrix} 1 & [a] \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & c \\ -1 & d \end{pmatrix} = \begin{pmatrix} \{a\} & c + [a]d \\ -1 & d \end{pmatrix}.$$

Cette identification s'interprète très bien en termes de domaines fondamentaux : on passe d'un domaine fondamental où le plus petit rectangle est de largeur 1 et le plus grand de largeur a à un autre où le plus grand rectangle est de largeur 1 et le plus petit de largeur $\{a\}$ (cf. Fig. 1). De la même façon, Λ_1 s'identifie à Σ_0 .

Il est alors facile de calculer l'application \tilde{T} sur Σ_0 : partant du point (x, y) de Σ_0 , le flot géodésique arrive au temps $-2 \log x$ au point $(1/x, xy)$ de Λ_0 , qui est identifié au point $(\{1/x\}, x - x^2y)$ de Σ_1 . Après un calcul symétrique pour les points de Σ_1 , on retrouve la formule donnée ci-dessus.

□

Remarque. Au lieu de paramétrer Σ_0 par (b, c) , on peut prendre comme coordonnées $(b, c/d)$, c'est-à-dire prendre le rapport des hauteurs et le rapport des largeurs (puisque sur Σ_0 on a $a = 1$, donc $b = b/a$) plutôt que la hauteur et la largeur du plus petit rectangle; ce système de coordonnées fait jouer un rôle symétrique aux largeurs et aux hauteurs, il envoie Σ_0 sur le carré unité, et un calcul simple montre que l'application de premier retour s'écrit alors comme un revêtement de l'application \bar{T}_1 décrite au paragraphe 1.

□

De la mesure invariante pour \tilde{T} on déduit une mesure invariante pour g_t , qui a diverses interprétations : mesure de Haar sur $SL(2, \mathbb{R})$, mesure de Liouville pour le flot géodésique, mesure donnée par la métrique riemannienne naturelle sur la surface modulaire... contrairement aux autres, la mesure induite par la métrique est complètement définie, et pas seulement à une constante multiplicative près, et on peut montrer que la mesure transverse invariante qu'elle définit sur $\tilde{\Sigma}$ est exactement la mesure de Lebesgue; en particulier, le volume de $\tilde{\Sigma}$ pour cette mesure transverse est $2 \log 2$; nous utilisons ces faits dans la prochaine section.

En utilisant l'application définie à la fin de la section 1, on peut facilement donner un codage explicite de l'application \tilde{T} , c'est-à-dire une conjugaison avec un revêtement d'ordre 2 du décalage S sur $\mathbb{N}^{\ast\mathbb{Z}}$. L'ensemble I défini à la fin de la section 1 correspond ici à l'intersection de la surface de section avec les géodésiques qui repassent une infinité de fois,

dans le passé et le futur, dans un domaine compact; le codage est évidemment plus facile dans ce domaine, mais il est aussi possible dans le complémentaire, au moyen de suites finies; nous n'insisterons pas sur ce sujet, sauf pour remarquer que des complications de ce type sont inévitables quand on essaye de conjuguer une application sur une surface et une application sur un ensemble de Cantor comme $\mathbb{N}^*\mathbb{Z}$.

On a même un résultat plus précis que le codage de l'application de premier retour, puisque tout élément du domaine fondamental s'écrit de façon unique $g_t(x, y, \epsilon)$, avec $(x, y, \epsilon) \in \tilde{\Sigma}$ et $0 \leq t < -2 \log x$; rappelons qu'étant données une bijection T sur un ensemble X et une application f définie sur X à valeurs réelles positives, telle que la série $\sum f(T^n x)$ tende vers l'infini pour tout point x de X , on appelle flot spécial au dessus de X , d'application de premier retour T et de temps de retour f , le flot ϕ_s sur $\{(x, t), x \in X, 0 \leq t < f(x)\}$ défini par $\phi_s(x, t) = (x, t + s)$ si $t + s < f(x)$, et $\phi_s(x, t) = (Tx, 0)$ si $s + t = f(x)$ (ces formules permettent de définir $\phi_s(x, t)$ pour tout s).

Corollaire 6.2. *L'application qui au point $g_t(x, y, \epsilon)$ associe $((a_n)_{n \in \mathbb{Z}}, \epsilon, t)$, où (a_n) est le codage de (x, y) défini dans la section 1, conjugue le flot géodésique à un flot spécial ψ_t au dessus de $\mathbb{N}^*\mathbb{Z} \times \{0, 1\}$, de temps de retour $-2 \log[0; a_1, \dots, a_n, \dots]$ et dont l'application de premier retour est un revêtement d'ordre 2 du décalage.*

Il est intéressant de remarquer que le temps de retour à la section ne dépend que de la coordonnée x (ou de la partie d'indices positifs de la suite (a_n) associée).

Si, au lieu de s'intéresser aux points individuels, on regarde les géodésiques, on obtient un codage par les suites d'entiers positifs :

Corollaire 6.3. *On a une bijection entre les géodésiques de la surface modulaire et les couples (u, ϵ) (u suite finie ou infinie d'entiers positifs, $\epsilon \in \{0, 1\}$) modulo l'équivalence $(u, \epsilon) \equiv (Su, 1 - \epsilon)$. En particulier, une géodésique tend vers l'infini si et seulement si la suite u n'est pas définie au-delà d'un certain rang, elle vient de l'infini si et seulement si elle n'est pas définie avant un certain rang. Une géodésique est périodique si et seulement si la suite (u_n) associée est périodique, et dans ce cas, si p est la plus petite période paire de la suite, la longueur de la géodésique est*

$$\sum_{j=0}^{p-1} -2 \log[0; \overline{u_{j+1}, \dots, u_{j+p}}].$$

Preuve. Seule la dernière assertion demande une démonstration; mais il suffit de calculer le temps nécessaire pour aller du point $(u, \epsilon, 0)$ à lui-même, en calculant les p temps de retour intermédiaire à la section, d'après la formule donnée au corollaire précédent.

□

7. La constante de Lévy et le volume du fibré tangent à la surface modulaire

Si l'on tronque à l'ordre n le développement en fraction continue d'un nombre

$$x = [0; a_1, \dots, a_n, \dots],$$

on obtient un nombre rationnel p_n/q_n , appelé convergent d'ordre n de x . Ces nombres se calculent facilement par récurrence, et on a les formules :

$$\begin{aligned} p_0 &= 0 & p_1 &= a_0 a_1 + 1 & p_n &= a_n p_{n-1} + p_{n-2} \text{ si } n \geq 2 \\ q_0 &= 1 & q_1 &= a_1 & q_n &= a_n q_{n-1} + q_{n-2} \text{ si } n \geq 2. \end{aligned}$$

Les convergents sont les meilleures approximations rationnelles de x , ils satisfont $|x - p_n/q_n| < 1/q_n^2$. Pour évaluer la vitesse d'approximation, il est intéressant de connaître la croissance des q_n ; celle-ci est donnée, pour presque tout nombre, par la proposition suivante, due à Lévy (cf. [Lev36]) :

Proposition 7.1. *Pour presque tout nombre, la suite des dénominateurs des convergents satisfait :*

$$\lim_{n \rightarrow \infty} \frac{\log q_n}{n} = \frac{\pi^2}{12 \log 2}$$

Preuve. Nous allons étudier la géodésique issue de

$$\begin{pmatrix} 1 & 0 \\ -x & 1 \end{pmatrix}$$

Notons $\tau_1, \tau_2, \dots, \tau_n, \dots$ ses temps successifs d'intersection avec la surface $\tilde{\Sigma}$; quitte à écarter un ensemble de mesure nulle, nous pouvons supposer qu'il y a une infinité d'intersections. On définit une suite x_n par $x_{-1} = 1$, $x_0 = x$, $x_{n+1} = x_{n-1} - a_{n+1}x_n$; on vérifie facilement que $x_n/x_{n-1} = T^n(x)$. On montre par récurrence que le point d'intersection d'ordre $2n$ avec Σ est donné par :

$$\begin{pmatrix} x_{2n-1} & q_{2n-1} \\ -x_{2n} & q_{2n} \end{pmatrix} \begin{pmatrix} e^{\frac{\tau_{2n}}{2}} & 0 \\ 0 & e^{-\frac{\tau_{2n}}{2}} \end{pmatrix}$$

et une formule analogue pour l'intersection d'ordre $2n + 1$; le point essentiel consiste à voir que l'on passe d'un point au suivant en multipliant à droite par une matrice diagonale correspondant au flot, et à gauche par une matrice entière correspondant à un changement de coordonnées; cette matrice est de la forme

$$\begin{pmatrix} 1 & 0 \\ a_{2n} & 1 \end{pmatrix} \quad \text{ou} \quad \begin{pmatrix} 1 & a_{2n+1} \\ 0 & 1 \end{pmatrix}$$

suivant la parité. □

On a le lemme suivant :

Lemme 7.2. *Soit $\begin{pmatrix} 1 & c \\ -b & d \end{pmatrix}$ un élément de Σ_0 ; on a toujours $1/2 \leq d \leq 1$.*

Preuve du lemme. On a $d + bc = 1$, b et c sont positifs, donc d est inférieur à 1; comme c est inférieur à d et b inférieur à 1, $2d$ est supérieur à 1, d'où le résultat. □

Suite de la preuve. Appliqué aux points calculés plus haut, ce lemme implique que, pour tout point x , on a : $1/2 \leq q_n e^{\tau_n/2} \leq 1$. Si l'orbite de x recoupe une infinité de fois la surface de section, en prenant le logarithme et en divisant par n , on en déduit :

$$\lim_{n \rightarrow \infty} \frac{\log q_n}{n} - \frac{\tau_n}{2n} = 0$$

Autrement dit, le terme qui apparaît dans le théorème de Lévy est la moitié du temps de retour moyen sur l'orbite. Compte tenu du fait que le temps de premier retour d'un point

$\begin{pmatrix} 1 & c \\ x & d \end{pmatrix}$ ne dépend que de x , et du codage donné au paragraphe précédent, on voit que, si l'on appelle $\tau(x)$ la fonction temps de premier retour, on a :

$$\frac{\tau_n}{n} = \frac{1}{n} \sum_{i=0}^{n-1} \tau(T^i x)$$

Cette expression est la somme de Birkhoff associée à la fonction temps de retour. Mais on sait que le flot géodésique sur la surface modulaire est ergodique; donc cette fonction tend presque partout vers une constante, qui est la moyenne du temps de premier retour sur la surface $\tilde{\Sigma}$. Cette moyenne est elle-même le quotient, par le volume de la surface, de l'intégrale de ce temps de retour sur $\tilde{\Sigma}$, qui n'est autre que le volume de l'espace tout entier. Cet espace est le fibré tangent à la surface modulaire. Cette surface est d'aire $\pi/3$, puisqu'elle admet dans le plan hyperbolique un domaine fondamental qui est un triangle isocèle d'angle $0, \pi/3, \pi/3$; il suffit pour obtenir l'aire d'appliquer la formule de Gauss pour les triangles hyperboliques. Le fibré tangent a pour fibre $PSO(2, \mathbb{R})$, qui est de longueur π (ne pas oublier que $-Id$ agit de façon triviale, c'est pour cela que la fibre a pour longueur π et non 2π comme on s'y attend); le volume total de l'espace est donc $\pi^2/3$.

On a vu plus haut que l'aire de $\tilde{\Sigma}$ est $2 \log 2$; le temps de retour moyen est donc $\pi^2/(6 \log 2)$, et compte tenu du facteur 2 introduit dans le calcul, on retrouve bien la constante cherchée.

□

8. Un codage du flot géodésique sur un alphabet fini

Un choix différent de la surface de section conduit à retrouver d'autres algorithmes de fractions continues. En particulier, si, dans la procédure donnée au paragraphe 4, on inverse les deux dernières étapes, c'est-à-dire si on prolonge le segment vertical jusqu'à rencontrer une deuxième fois le segment horizontal avant de poursuivre celui-ci (cf. Fig. 6), on obtient un domaine fondamental différent, caractérisé par $0 \leq a \leq 1 \leq b \leq 1 + a$, sans condition sur les deux autres coefficients, sauf $ad + bc = 1$.

On peut alors refaire la même étude qu'au paragraphe 5, en définissant deux domaines Ω'_0 et Ω'_1 , avec des faces rentrantes et sortantes. L'une des faces rentrantes, Σ'_0 , est formée des matrices du type

$$\begin{pmatrix} a & c \\ -1 & d \end{pmatrix}$$

avec $a < 1$ et c, d positifs, et Σ'_1 est définie de façon symétrique. La surface de section sur laquelle on définit l'application de premier retour est $\Sigma'_0 \cup \Sigma'_1$.

On peut paramétrer Σ'_0 par a et c , tous deux compris entre 0 et 1, et $\tilde{\Sigma}'$ par (a, c, ϵ) , avec $\epsilon = 0$ ou 1. Un calcul simple montre que l'application de premier retour en $\tilde{\Sigma}'$ est donnée par

$$\begin{aligned} \tilde{T}'_1 : \tilde{\Sigma}' \rightarrow \tilde{\Sigma}', \quad (a, c, \epsilon) &\mapsto \left(\frac{a}{1-a}, c(1-a), \epsilon \right) \text{ si } a < \frac{1}{2} \\ (a, c, \epsilon) &\mapsto \left(\frac{1-a}{a}, 1-c(1-a), 1-\epsilon \right) \text{ si } a \geq \frac{1}{2}. \end{aligned}$$

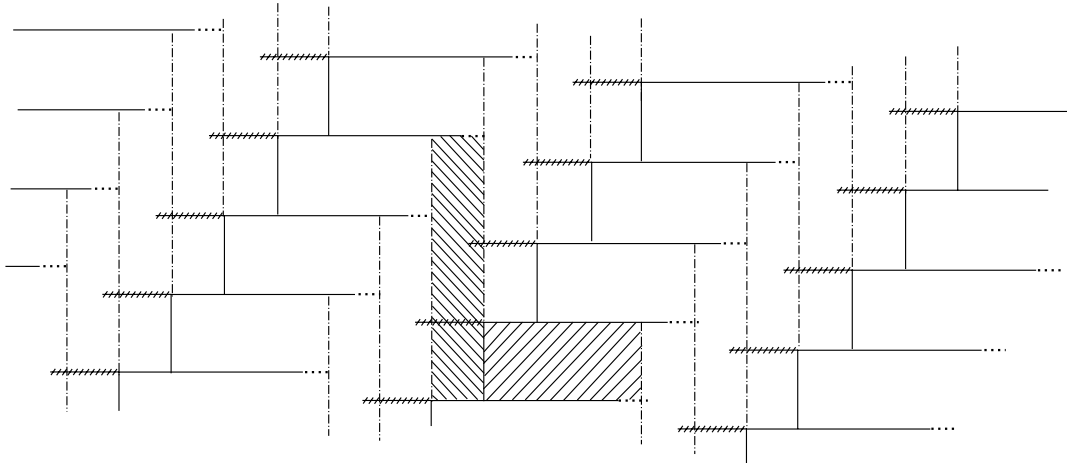


FIGURE 6.

C'est un revêtement d'ordre 2 de l'application représentée ci-dessous (cf. Fig. 7); elle a un domaine de définition très simple, mais des formules un peu compliquées (en particulier, elle ne préserve pas la mesure de Lebesgue). En utilisant la partition de Markov naturelle pour cette fonction, on peut coder le flot géodésique par un alphabet à deux lettres.

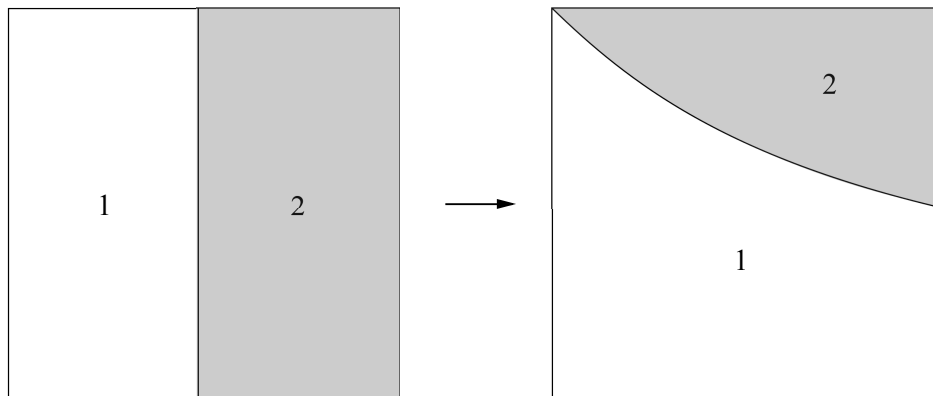


FIGURE 7.

En utilisant comme coordonnées non plus (a, c) mais (a, d) , avec $0 \leq a < 1$ et $0 \leq d < 1/a$, on trouve pour l'application de premier retour la forme suivante :

$$\begin{aligned} \tilde{T}' : \tilde{\Sigma}' \rightarrow \tilde{\Sigma}', \quad & (a, d, \epsilon) \mapsto \left(\frac{a}{1-a}, d(1-a)^2 + (1-a), \epsilon \right) \text{ si } a < \frac{1}{2} \\ & (a, d, \epsilon) \mapsto \left(\frac{1-a}{a}, a - a^2d, 1 - \epsilon \right) \text{ si } a \geq \frac{1}{2}. \end{aligned}$$

C'est un revêtement de l'application schématisée dans la figure 8; cette application préserve la mesure de Lebesgue, et son domaine est de mesure infinie.

Dans les deux cas, l'application de premier retour se projette sur la première coordonnée en l'application T' de l'intervalle $[0, 1]$ définie par $T'(x) = x/(1-x)$ si $x < 1/2$, $T'(x) = (1 -$

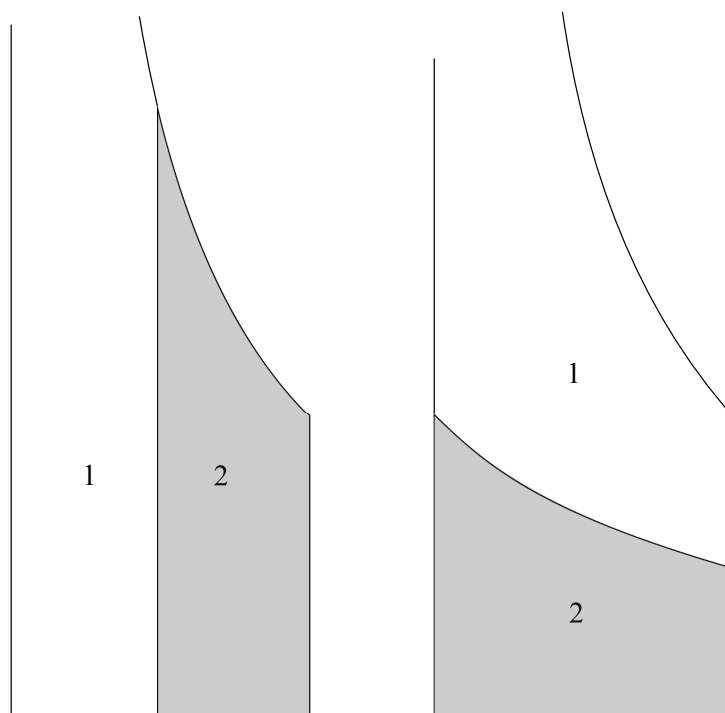


FIGURE 8.

$x)/x$ si $x \geq 1/2$. On peut voir cette application comme la version projective de l'algorithme qui consiste à prendre deux nombres positifs a et b , à soustraire le plus petit du plus grand et à itérer l'opération; on réalise ainsi l'algorithme d'Euclide en ne faisant que des soustractions, donc avec beaucoup d'étapes intermédiaires. Il s'agit d'une version « lente » de l'algorithme des fractions continues : si x est compris entre $1/(n+1)$ et $1/n$, alors $\{1/x\} = T^n(x)$.

RÉFÉRENCES

- [AF91] R. Adler and L. Flatto. Geodesic flows, interval maps and symbolic dynamics. *Bull. Amer. Math. Soc.*, 25 :229–334, 1991.
- [AN] P. Arnoux and A. Nogueira. Mesures de gauss pour des algorithmes de fractions continues multidimensionnelles (à paraître aux Annales Scientifiques de l'Ecole Normale supérieure).
- [Art24] E. Artin. Ein mechanisches System mit quasiergodischen Bahnen. *Abh. Math. Sem. Univ. v. Hamburg*, 3 :170–175, 1924.
- [Ito84] S. Ito. Number theoretic expansions, algorithms and metrical observations. *Séminaire de théorie des nombres de Bordeaux*, pages 3.01–3.27, 1984.
- [Lev36] P. Levy. Sur le développement en fraction continue d'un nombre choisi au hasard. *Composition Mathematica*, 3 :286–303, 1936.
- [Nak81] H. Nakada. Metrical theory for a class of continued fraction transformations and their natural extensions. *Tokyo J. Math.*, 4 :399–426, 1981.
- [Roh64] V. A. Rohlin. Exact endomorphisms of Lebesgue spaces. *Amer. Math. Soc. Transl.*, 39 :1–36, 1964.
- [Ser85] C. Series. The modular surface and continued fractions. *J. London Math. Soc.*, 31 :69–80, 1985.
- [Vee84] W.A. Veech. The metric theory of interval exchange transformations I,II,III. *Amer. J. Math.*, 106 :1331–1422, 1984.
- [Vee86] W.A. Veech. The Teichmüller geodesic flow. *Annals of Math.*, 124 :441–530, 1986.