Ergod. Th. & Dynam. Sys. (1984), **4**, 487-492
Printed in Great Britain

The backward continued fraction map and geodesic flow

ROY L. ADLER AND LEOPOLD FLATTO

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, *USA*;
Bell Laboratories, Murray Hill, New Jersey 07974, *USA*

(*Received* 1 *August* 1983)

Abstract. The 'backward continued fraction' map studied by A. Reyni is defined by $y = g(x)$ where $g(x)$ equals the fractional part of $1/(1-x)$ for $0<x<1$. We show that it is a factor map of a special cross-section map for the geodesic flow on the unit tangent bundle of the modular surface. This gives an alternative derivation of the fact that this map preserves the infinite measure dx/x on the unit interval.

1. *Introduction*

In recent years much work has been done on the iterates of maps of the unit interval into itself. (An extensive reference list is given in [**4**].) The subject centres mainly on two problems: (i) the ergodic properties of the map; (ii) the existence of an invariant measure (which we assume to be equivalent to Lebesgue measure). Concerning (ii), most results are merely of an existential nature. I.e. one proves that under certain conditions there is a unique invariant measure and one investigates its smoothness properties. Only in rare cases is a formula produced. One of these rare cases is given by the continued fraction transformation

$$y = f(x) = \left(\frac{1}{x}\right), \qquad 0<x<1.$$

The reason for the name stems from the fact that $f(a_1a_2\cdots a_n\cdots) = a_2a_3\cdots a_n\cdots$, where $a_1a_2\cdots a_n\cdots$ is the continued fraction expansion of x. For this map, Gauss has shown that $dx/1+x$ is the invariant measure. Using this and the ergodic properties of $f(x)$, it is possible to conclude from the ergodic theorem some interesting facts concerning the frequency of digits appearing in the continued fraction expansion of real numbers [**3**].

Another case where an explicit formula is obtainable, probably not as well known as the previous one, is the map

$$y = g(x) = \left(\frac{1}{1-x}\right), \qquad 0<x<1.$$

The graph of $g(x)$ is obtained from that of $f(x)$ by flipping the latter about the vertical line $x=\frac{1}{2}$, (figure 1). For this reason, we call $y = g(x)$ the backward continued fraction transformation. Here the invariant measure is dx/x. This formula, apparently attributed to Renyi [**5**], is derived as follows. For $0<t<1$,

$$g^{-1}(t, 1) = \bigcup_{n=1}^{\infty} [1-(n+1)^{-1}, 1-(n+t)^{-1}].$$

Hence

$$\int_{g^{-1}(t,1)} \frac{dx}{x} = \sum_{n=1}^{\infty} \int_{1-(n+t)^{-1}}^{1-(n+1)^{-1}} \frac{dx}{x} = \log \prod_{n=1}^{\infty} \frac{n}{n+1} \frac{n+t}{n+t-1}$$
$$= -\log t = \int_t^1 \frac{dx}{x}.$$

Observe that $\int_0^1 dx/(1+x) < \infty$, $\int_0^1 dx/x = \infty$, so that the two transformations $y = f(x)$, $y = g(x)$ are not conjugate to each other by a non-singular measurable change of variables.

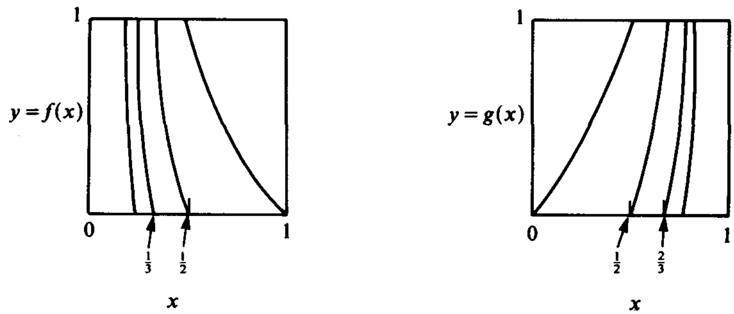

FIGURE 1

In [1] we found a relation between the ergodic properties of maps of the unit interval and the ergodic properties of geodesic flows on two dimensional surfaces of constant negative curvature. In particular, we studied the modular surface and showed, by simple geometric arguments, that the continued fraction map $f(x)$ is a factor of a certain cross section map associated with the geodesic flow on this surface. From this fact it is possible to derive anew the invariant measure for $f(x)$ from the invariant hyperbolic measure associated with the geodesic flow.

In the present note we show that the same can be done for the backward continued fraction map. To make the paper self contained, we describe in § 2 the modular surface and its geodesic flow. In § 3, we describe the cross section and the related cross section map. Finally, in § 4 we show how the backward continued fraction map $y = g(x)$ arises as a factor of the cross section map and derive from this the invariant measure of $g(x)$.

2. *Geodesic flow*

We describe the geodesic flow on the modular surface. It proves convenient to first describe the corresponding flow on the hyperbolic plane.

Let $\mathbf{H} = \{x + iy: y > 0\}$ be the hyperbolic plane. The metric on $\mathbf{H}$ is given by $ds^2 = (dx^2 + dy^2)/y^2$ and the geodesics for it are the half circles and straight lines orthogonal to the x-axis. Let $\mathbf{U}$ be the unit tangent bundle consisting of unit tangent vectors on $\mathbf{H}$. $\mathbf{U}$ is coordinatized by $u = u(x, y, \theta)$, where (x, y) is the base point of $u \in \mathbf{U}$ and θ is the angle measured counterclockwise between the positive x-axis and u. The geodesic flow G_t, $-\infty < t < \infty$, is the class of homeomorphisms of $\mathbf{U}$

https://doi.org/10.1017/S0143385700002583 Published online by Cambridge University Press

defined by $u \to u_t$, where u and u_t are the initial and terminal unit tangent vectors of a geodesic segment of length t. G_t has a simple description if we use the following coordinates. To each $u \in \mathbf{U}$ assign ξ, η, s where ξ, η are the points on the x-axis of the geodesic γ determined by u, ξ being the point in the forward direction, and s is the hyperbolic distance on γ measured from some conveniently chosen origin. In these coordinates we have

$$G_t: (\xi, \eta, s) \to (\xi, \eta, s+t). \tag{2.1}$$

The hyperbolic measure is given by

$$dm = \frac{d\xi\, d\eta\, ds}{(\xi-\eta)^2}. \tag{2.2}$$

From (2.1), (2.2) we see that dm is invariant under G_t.

The above concepts carry over to the modular surface. Let Γ be the modular group

$$\Gamma = \left\{ \tau(z) = \frac{az+b}{cz+d} : ad-bc=1;\, a, b, c, d \in \mathbb{Z} \right\}.$$

Γ acts both on $\mathbf{H}$ and $\mathbf{U}$, the action on the latter being denoted by

$$\bar{\tau}(z, \theta) = (\tau(z), \theta + \arg \tau'(z))$$

where we have written $(z, \theta) = (x+iy, \theta)$ for (x, y, θ). We refer to Γ as $\bar{\Gamma}$ when acting on $\mathbf{U}$. Let M, $\mathcal{M}$ be respectively the spaces of Γ- and $\bar{\Gamma}$-orbits on $\mathbf{H}$, $\mathbf{U}$. To obtain concrete realizations of M, $\mathcal{M}$ we introduce

$$F = \{z = x+iy : |z| > 1, -\tfrac{1}{2} < x < \tfrac{1}{2}\}.$$

F is a fundamental domain for Γ, which means that:

(i) $\tau_1 F \cap \tau_2 F = \emptyset$ for any two distinct elements $\tau_1, \tau_2 \in \Gamma$; and

(ii) $\mathbf{H} = \bigcup_{\tau\in\Gamma} \tau\bar{F}$, $\bar{F}$ being the closure of F.

Thus $\mathbf{H}$ is tesselated with the images of F under Γ (see figure 2).

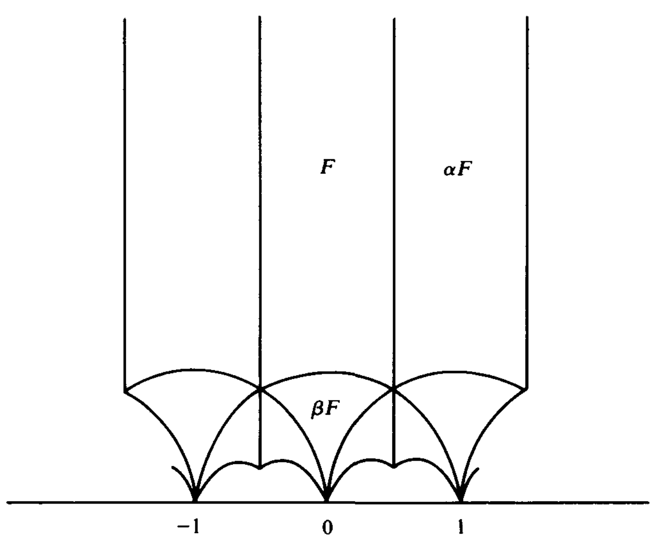

FIGURE 2

https://doi.org/10.1017/S0143385700002583 Published online by Cambridge University Press

The transformations $\alpha(z)=z+1$, $\beta(z)=-1/z$ generate Γ. Opposite vertical boundary lines of F are identified under α, and the left half of the bottom boundary with the right half under β. Consequently M can be thought of as $\bar{F}$ under these identifications. Similarly, unit vectors with base point in the boundary of F can also be identified under $\bar{\alpha}$ and $\bar{\beta}$, and so $\mathcal{M}$ can be thought of as unit vectors emanating from points of $\bar{F}$ under these identifications.

To coordinatize M and $\mathcal{M}$, we introduce the projection maps

$$\pi(z)=\Gamma z, \qquad z\in \mathbf{H};$$
$$\bar{\pi}(u)=\bar{\Gamma} u, \qquad u\in \mathbf{U};$$

where Γz, $\bar{\Gamma} u$ denote the Γ- and $\bar{\Gamma}$-orbits of z and u. π and $\bar{\pi}$ map respectively $\mathbf{H}$ onto M and $\mathbf{U}$ onto $\mathcal{M}$. With the exception of the Γ-orbits of the points $i, \frac{1}{2}+(\sqrt{3}/2)i$ and the $\bar{\Gamma}$-orbits of the unit vectors based at $i, \frac{1}{2}+(\sqrt{3}/2)i$, π and $\bar{\pi}$ are locally 1–1, and so x, y or ξ, η provide local coordinates at points of $M\setminus\{\pi(i), \pi(\frac{1}{2}+(\sqrt{3}/2)i)\}$, and x, y, θ or ξ, η, s provide local coordinates at points of $\mathcal{M}\setminus\{\bar{\pi}(0,1,\theta), \bar{\pi}(\frac{1}{2}, \sqrt{3}/2, \theta), 0\le\theta<2\pi\}$. If we avoid the exceptional points then the formulae for ds, dm carry over to M and $\mathcal{M}$. Because G_t commutes with $\bar{\Gamma}$ on U, $\mathcal{M}$ inherits the geodesic flow $\bar{G}_t=\bar{\pi}G_t\bar{\pi}^{-1}$. Because m is invariant under the elements of $\bar{\Gamma}$, $\bar{G}_t$ has the invariant measure $\bar{m}$ which is defined as the m measure of any local inverse of $\bar{\pi}$.

3. *Cross section map*

A cross section on $\mathcal{M}$ is a subset of $\mathcal{M}$ which every $\bar{G}_t$-orbit meets infinitely often, both past and future. The correspondence between successive return points serves to define the cross section map. The cross section C which we choose consists of the $\bar{\pi}$-projections of those $u\in\mathbf{U}$ with base point in Y^+, the positive y-axis, and pointing to the right (see figure 3). We observe that the chosen elements are all distinct. For if u has base point in $Y^+\cap\beta F$ and points to the right, then $\bar{\beta}u$ has base point in $Y^+\cap F$ and points to the left. As given, C does not quite meet the requirements of a cross section, as there are $\bar{G}_t$-orbits which do not visit it infinitely both past and future – these are the $\bar{\pi}$-projections of G_t-orbits starting or terminating at cusp points. Analytically, these $\bar{G}_t$-orbits are described by $\bar{\pi}(\xi, \eta, s)$, $-\infty<s<\infty$, with either ξ or η rational or ∞. To get rid of this difficulty, we remove these points, which comprise a set of measure zero, from $\mathcal{M}$. Thus we tacitly assume from now on that all ξ, η in consideration are irrational.

We assign to each $\bar{u}=\bar{\pi}(u)\in C$ its ξ, η coordinates. In this coordinate description

$$C=\{(\xi,\eta): \xi>0, \eta<0\}.$$

To describe the cross section map T_C, we decompose C as $C_1\cup C_2$ where

$$C_1=\{(\xi,\eta): 0<\xi<1, \eta<0\}, \qquad C_2=\{(\xi,\eta): \xi>1, \eta<0\}.$$

To obtain an expression for T_C consider the hyperbolic triangle $\triangle$ with vertices $0, 1, \infty$ depicted in figure 3. The three sides of $\triangle$ are equivalent under Γ, the sides $\widehat{01}$, $\widehat{1\infty}$ being carried respectively into $\widehat{0\infty}$ by the transformations $w=z/1-z$, $w=z-1$. Furthermore, Γ-orbits of interior points of Δ are distinct from those of boundary

https://doi.org/10.1017/S0143385700002583 Published online by Cambridge University Press

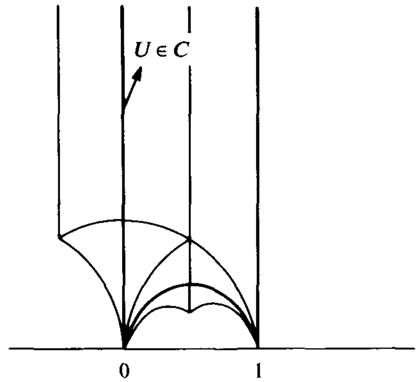

FIGURE 3

points. Let u be as above and let γ be the geodesic determined by u. If $\bar{\pi}(u) \in C_2$, then γ leaves $\triangle$ through the right vertical wall. Let the unit tangent at the point of departure be (ξ, η, s). γ is identified under $w = z - 1$ with a geodesic entering $\triangle$ at the left vertical wall, the unit tangent at the point of entrance being $(\xi - 1, \eta - 1, s)$. Hence in this case

$$T_C(\xi, \eta) = (\xi - 1, \eta - 1).$$

If $\bar{\pi}(u) \in C_1$, then γ leaves $\triangle$ through the base $\widehat{01}$ with unit tangent (ξ, η, s). γ is identified under $w = z/1 - z$ with a geodesic entering $\triangle$ at the left vertical wall with unit tangent $(\xi/1 - \xi, \eta/1 - \eta, s)$. Hence in this case

$$T_C(\xi, \eta) = \left(\frac{\xi}{1-\xi}, \frac{\eta}{1-\eta}\right).$$

Thus

$$T_C(\xi, \eta) = \begin{cases} \left(\dfrac{\xi}{1-\xi}, \dfrac{\eta}{1-\eta}\right) & \text{on } C_1 \\ (\xi - 1, \eta - 1) & \text{on } C_2. \end{cases} \tag{3.1}$$

The sets C_i and their images $C_i' = T_C(C_i)$ are depicted in figure 4 where C_i, C_i' are replaced respectively by i, i'.

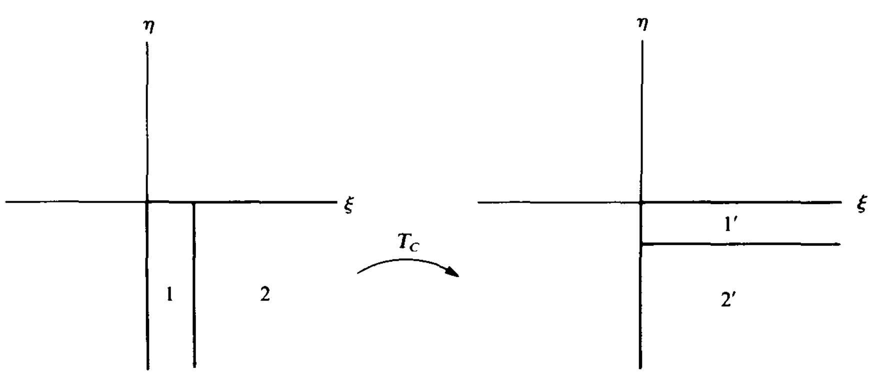

FIGURE 4

https://doi.org/10.1017/S0143385700002583 Published online by Cambridge University Press

As explained in [1], the invariant measure dm_C for T_C is obtained from dm by dropping ds, i.e.

$$dm_C = \frac{d\xi d\eta}{(\xi-\eta)^2}. \tag{3.2}$$

4. *Factor map*

From (3.1) we see that $G(\xi)$ is a factor map of $T_C(\xi, \eta)$ where

$$G(\xi) = \begin{cases} \dfrac{\xi}{1-\xi}, & 0<\xi<1 \\ \xi - 1, & 1<\xi<\infty. \end{cases} \tag{4.1}$$

The invariant measure $h(\xi)\, d\xi$ for $G(\xi)$ is obtained by integrating dm_C with respect to η [1]. I.e.

$$h(\xi) = \int_{-\infty}^{0} \frac{d\eta}{(\xi-\eta)^2} = \frac{1}{\xi}, \qquad \xi > 0. \tag{4.2}$$

The backward continued fraction map $g(\xi)$ is $G(\xi)$ induced to $(0, 1)$. I.e. for $0<\xi<1$, let $n(\xi)$ be the smallest positive integer such that $G^n(\xi)$, the n'th iterate of $G(\xi)$, is in $(0, 1)$. Then

$$g(\xi) = G^{n(\xi)}(\xi) = \left(\frac{\xi}{1-\xi}\right) = \left(\frac{1}{1-\xi}\right), \qquad 0<\xi<1. \tag{4.3}$$

It follows that the invariant measure for $g(\xi)$ is also $d\xi/\xi$ [2].

Finally, we remark that $g(\xi)$ is itself a factor of a cross section map. We just interchange the operations of factoring and inducing given above. Let

$$Q = \{(\xi, \eta) : 0<\xi<1, \eta<0\}$$

and let $T_Q(\xi, \eta)$ be $T_C(\xi, \eta)$ induced to Q. It follows from (3.1) that

$$T_Q(\xi, \eta) = \left(\left(\frac{\xi}{1-\xi}\right), \frac{\eta}{1-\eta} - \left[\frac{\xi}{1-\xi}\right]\right), \qquad (\xi, \eta) \in Q. \tag{4.4}$$

$T_Q(\xi, \eta)$ is the cross section map for the cross section Q and has the invariant measure $d\xi\, d\eta/(\xi-\eta)^2$. (4.4) shows that $T_Q(\xi, \eta)$ has the factor $g(\xi)$ with the invariant measure $d\xi/\xi$.

REFERENCES

[1] R. L. Adler & L. Flatto. Cross section map for the geodesic flow on the modular surface. In *Conference in Modern Analysis and Probability* (Contemporary Math. **26** (1984), 9-24). Amer. Math. Soc: Providence, R.I.

[2] R. L. Adler & B. Weiss. The ergodic infinite measure preserving transformation of Boole. *Israel J. of Math.* **16** (1973), 263–278.

[3] P. Billingsley. *Ergodic Theory and Information.* John Wiley Sons, Inc.: New York (1965).

[4] Z. Nitecki. Topological dynamics on the interval. In *Ergodic Theory and Dynamical Systems,* (Proc. Special Year Md. 1979–1980) Progress in Math., vol. 2, 1–73. Birkhauser: Boston, Basel & Stuttgart.

[5] A. Renyi. Valós számok elöállitására szolgáló algoritmusokról. *M.T.A. Mat. és Fiz. Oszt. Közl.* **7** (1957), 265–293.