LITERATURE CITED

1. Ven'-tuan Lu (Wen-duan Lu), "On imbedding theorems for spaces of functions with partial derivatives, summable with various exponents," Vestnik Leningrad. Univ. Mat. Mekh. Astronom., 16, No. 7, 23-37 (1961).
2. S. N. Kruzhkov, "Boundary value problems for second-order degenerate elliptic equations," Mat. Sb., 77, No. 3, 299-334 (1968).
3. S. N. Kruzhkov and I. M. Kolodii, "On the theory of anisotropic Sobolev spaces," Usp. Mat. Nauk, 38, No. 2, 207-208 (1983).
4. S. N. Kruzhkov and A. G. Korolev, "On the imbedding theory of anisotropic function spaces," Dokl. Akad. Nauk SSSR, 285, No. 5, 1054-1057 (1985).
5. S. L. Sobolev, "On estimates of certain sums for functions defined on a net," Izv. Akad. Nauk SSSR, Ser. Mat., 4, No. 1, 5-16 (1940).
6. Yu. I. Mokin, "A net analogue of the imbedding theorem for classes of type W," Zh. Vychisl. Mat. i Mat. Fiz., 11, No. 6, 1361-1373 (1971).

ERGODICITY OF TRANSITIVE UNIMODAL TRANSFORMATIONS OF A SEGMENT

A. M. Blokh and M. Yu. Lyubich

UDC 517.1

1. Introduction

We consider a piecewise-monotonic transformation f of the segment $I = [0, 1]$ with negative Schwarzian. This means that the function $f \in C^3$ has a finite number of critical points outside which $Sf = f'''/f' - (3/2)(f''/f')^2 < 0$. The transformation f is called unimodal if it has a unique critical point c while this point is nondegenerate. Despite the elementary nature of the situation, such a transformation generates a nontrivial and interesting dynamical system [1, 2].

The transformation f is called (topologically) transitive if it has a dense orbit $\{f^n x\}_{n=0}^{\infty}$ (where f^n is the n-th iteration of f) and ergodic if there does not exist a partition of $I = X_1 \cup X_2$ into two measurable invariant subsets of positive measure (invariance means that $fX_i \subset X_i$). Here the segment I is also called transitive (respectively, ergodic). The goal of the present paper is the proof of the following result.

THEOREM. A transitive unimodal transformation $f:I \to I$ with negative Schwarzian is ergodic.

We note two consequences of this theorem. The first has a well-known analog in the theory of rational endomorphisms of the Riemann sphere (Sullivan [3]). The set X is called strongly wandering if $f^n X \cap f^m X = \emptyset$ for $> m \geqslant 0$.

COROLLARY 1. A unimodal transformation f with regative Schwarzian has no strongly wandering sets X of positive measure for which all the iterations $f^n|X$ are injective.

COROLLARY 2. A unimodal transformation of a segment with negative Schwarzian can have at most one absolutely continuous invariant probability measure. If such a measure exists, then it is ergodic.

The basic result of the present paper was announced in [4].

2. Distortion Theorem for Functions with Negative Schwarzian

We denote by λ the Lebesgue measure on the line. A basic analytic instrument in the proof of the theorem is the following property of functions with negative Schwarzian.

Distortion Theorem [4, 5]. Let the map $\varphi : I \to J$ have no critical points inside I and $S\varphi < 0$. Let the interval $\mathcal{J}$ be divided into intervals $\mathcal{J}_1$ and $\mathcal{J}_2$ by the point y; let E be a measurable subset of $\mathcal{J}$, $\eta \in (0, 1)$. Then there exists an interval $K_\eta = [y, z_\eta]$ contained in some

---

Leningrad Section of the Mathematics Institute, Academy of Sciences of the USSR. Translated from Ukrainskii Matematicheskii Zhurnal, Vol. 41, No. 7, pp. 985-988, July, 1989. Original article submitted September 16, 1986.

0041-5995/89/4107-0841$12.50

$\mathcal{I}_l$, $l = 1, 2$, for which $\frac{\lambda(\varphi^{-1}(E \cap K_\eta))}{\lambda(\varphi^{-1}K_\eta)} \geqslant \eta \frac{\lambda(E \cap \mathcal{I}_l)}{\lambda(\mathcal{I}_l)}$.

For completeness we give a proof of the distortion theorem (different from the one proposed in [5]). It uses the famous Minimum Principle [1]. Let the map $\varphi: I \to J$ have no critical points inside I and $S\varphi < 0$. Then the function $|\varphi'|$ has no local minimum points inside I.

Proof of the Distortion Theorem. Let $x = \varphi^{-1}y$, I_1 and I_2 be the intervals into which the point x divides the segment I. By the minimum principle the fucntion $|\varphi'|$ is monotonically decreasing on one of these intervals, say on $I_1 = \varphi^{-1}\mathcal{I}_1$. After an affine change of coordinates one can assume that $I_1 = \mathcal{I}_1 = [0, 1]$, $x = y = 0$. Then $\varphi^{-1}K_\eta = (0, \alpha)$ and the inequality required assumes the form

$$\frac{1}{\alpha}\int_0^\alpha \chi_F dx \geqslant \eta \int_0^1 \chi_F |\varphi'| dx,$$

where χ_F is the characteristic function of the set $F = \varphi^{-1}E$. The following lemma completes the proof.

LEMMA 1. Let g be a continuous nonincreasing nonnegative function on the segment [0, 1], $\int_0^1 g(x)\,dx = 1$; $\psi \in L_1[0, 1]$. Then

$$\sup_{\alpha\in(0,1)} \frac{1}{\alpha}\int_0^\alpha \psi dx \geqslant \int_0^1 \psi g dx.$$

Proof. Setting $g_\alpha = \frac{1}{\alpha}\chi_{[0,\alpha]}$, we rewrite the inequality in the form

$$\sup_{\alpha\in(0,1)} \int_0^1 \psi g_\alpha dx \geqslant \int_0^1 \psi g dx.$$

Now it has become obvious since any function g of the class considered can be approximated uniformly by convex combinations of functions g_α.

We need some information about the topological properties of the dynamics of transformations of a segment [6]. A continuous transformation $f: [a, b] \to [a, b]$ is called mixing if for any interval $\mathcal{I} \subset [a, b]$ and any $\varepsilon > 0$ there exists an N such that $f^n\mathcal{I} \supset [a+\varepsilon, b-\varepsilon]$ for $n \geqslant N$. Clearly in this case for no n do there exist nontrivial f^n-invariant intervals.

LEMMA 2 [6]. Let $f: I \to I$ be a continuous transitive transformation of a segment. Then either f is mixing or there exists a fixed point a which divides I into two intervals I_1 and I_2 such that $f: I_1 \to I_2 \to I_1$ and the transformation $f^2: I_l \to I_l$ is mixing.

Further, we return to unimodal transformations. Let us assume, to be specific, that c is a maximum point of the function f. It follows easily from the transitivity of f that $f: c \mapsto 1 \mapsto 0$. Consequently, there exists a point $\gamma \in (c, 1]$, for which $f(\gamma) = f(0)$ and on the segment $[0, \gamma]$ there is defined an involution $\tau: x \mapsto x'$, where $f(x') = f(x)$. The smoothness of τ follows from the nondegeneracy of the critical point c.

In what follows we shall denote by (a, b) the interval with ends a and b without assuming that a < b. We also introduce the following notation: $U_b = (b, b')$, where $b \in [0, \gamma]$; $H_n(x)$ is the interval of monotonicity of the function f^n containing the point x; $M_n(x) = f^n$

LEMMA 3. Let f be a mixing unimodal transformation. Let us assume that the orbit of the point x enters the neighborhood U_b and n is the first time for which $f^n x \in U_b$. Then $M_n(x) \supset U_b$.

Proof. Let the point x divide the interval $H_n(x)$ into intervals H_n^+ and H_n^- the point c divide the interval U_b into intervals U_b^+ and U_b^-. Then for any $\mu = \pm 1$ and suitable $\gamma \in \{\pm 1\}$, $l < n$ we have $f^l H_n^\mu = [c, f^l x] \supset U_b^\gamma$. If $f^{n-l} c \in U_b$, then $f^{n-l}[c, f^l x] \subset U_b$ and all the more $f^{n-l}U_b = f^{n-l}U_b^\gamma \subset U_b$. But the last is impossible for mixing maps. Thus, the ends of the segment $M_n(x)$ are outside the limits of the neighborhood U_b which is what was needed.

We shall call a set symmetric, if it is τ-invariant.

Basic Lemma. Let $f: I \to I$ be a transitive unimodal transformation of a segment. Let the set $X \subset I$ have positive measure, be f-invariant and symmetric in a neighborhood of the critical point c. Then c is a point of condensation of the set X.

Proof. Let us assume that the transformation f is mixing (this does not restrict the generality by virtue of Lemma 2). By Guckenheimer's theorem [7], $\omega(x) \ni$ for a.a. $x \in I^*$ (cf. also [4, 5]). We fix a point of condensation $x \in X$ of the set X for which $\omega(x) \ni c$. Let $Y = I \setminus X$. For the interval $\mathcal{I} = (a, b)$ we denote by $\rho(a, b) = \rho(\mathcal{I}) = \lambda(Y \cap \mathcal{I})/\lambda(\mathcal{I})$ the density of the set Y in $\mathcal{I}$. We consider two cases:

1. The lower density of the set Y at the point c is positive; $\underline{\lim}_{b \to c} \rho(c-b, c+b) > 0$. Then by virtue of the smoothness of τ and the symmetry of the set X we have $\rho(U_b) \geqslant \varepsilon > 0$ for all $b \neq c$. One can assume that the point x has a symmetric one. Let $n_0 = 0$, n_{k+1} be the first time the trajectory $\{f^n x\}_{n=n_k+1}^{\infty}$ lands in the neighborhood $U_k \equiv U_{f^{n_k}x}$; x_k be the point of $f^{n_k}x$, $\tau(f^{n_k}x)$, which is located to the left of c. Since $U_k = \dot{\bigcup\limits_{i=k}^{\infty}} (U_i \setminus U_{i+1})$, for some sequence $k_j \to \infty$ one has $\rho(U_{k_j} \setminus U_{k_j+1}) \geqslant \varepsilon > 0$. Again using the symmetry of X and the smoothness of τ we get $\rho(x_{k_j}, x_{k_j+1}) \geqslant L^{-2}\rho(U_{k_j} \setminus U_{k_j+1}) \geqslant \varepsilon_1 > 0$, where L is the Lipschitz constant of the involution τ, $\varepsilon_1 = L^{-2}\varepsilon$. Moreover, for all k one has $\rho(x_k', x_{k+1}) \geqslant \min(\rho(x_k', c), \rho(c, x_{k+1})) \geqslant \varepsilon_1 > 0$. Thus, in each of the intervals into which the point x_{k_j+1} divides the interval U_{k_j} the density of the set Y is not less than ε_1.

We set $V_k = f^{-n_k} U_k \cap H_{n_k}(x)$. By Lemma 3, f^{n_k} maps V_k monotonically onto U_k. Applying the distortion theorem we get that in some half-neighborhood of the point x contained in V_k the density of the set Y is not less than $\varepsilon_1/2$. The transitivity of f implies the absence of homtervals†; consequently $\lambda(V_k) \to 0$. Thus, the set $y = I \setminus X$ has positive upper density at the point x, despite the fact that x is a point of condensation of the set X.

2. The lower density of the set Y at the point c is equal to 0: $\underline{\lim}_{b \to c} \rho(c-b,$ Arguing by contradiction, let us assume that $\overline{\lim}_{b \to c} \rho(c-b, c+b) > 0$. Then by virtue of the symmetry of X and the smoothness of τ one can find sequences $\{a_k\}$, $\{b_k\}$, such that

$$b_k \in (a_k, c), \; a_k \to c, \tag{1}$$

$$\rho(a_k, c) \geqslant \varepsilon, \; \rho(a_k', c) \geqslant \varepsilon > 0, \tag{2}$$

$$\rho(b_k, c) < \delta_k, \; \rho(b_k', c) < \delta_k, \quad \delta_k \to 0, \tag{3}$$

$$\rho(U_d) \leqslant \rho(U_{a_k}), \quad d \in U_{a_k} \setminus U_{b_k}. \tag{4}$$

Let n_k be the first time that the orbit $\{f^n x\}_{n=0}^{\infty}$ lands in the neighborhood U_{a_k}, $x_k = f^{n_k}x$. Replacing the points a_k, b_k by a_k', b_k', if necessary, we shall assume that $x_k \in (a_k, c)$. We have

$$\rho(x_k, a_k') \geqslant \frac{\lambda(Y \cap (c, a_k'))}{|a_k - a_k'|} \geqslant \frac{1}{1+L^2}\, \rho(U_{a_k}) \geqslant \frac{\varepsilon}{1+L^2} > 0.$$

In order to estimate the density of the set Y in the interval (a_k, x_k), we consider two cases:

a) $x_k \in (a_k, b_k]$. By virtue of (4), $\rho(U_{x_k}) \leqslant \rho(U_{a_k})$ and consequently, $\rho((x_k, a_k) \cup \tau(x_k, a_k)) = \rho(U_{a_k} \setminus U_{x_k}) \geqslant \rho(U_{a_k}) \geqslant \varepsilon$. Again using the smoothness of τ and the symmetry of X, we find $\rho(x_k, a_k) \geqslant L^{-2}\varepsilon$.

b) $x_k \in (b_k, c)$. Then $\lambda(Y \cap (a_k, c)) = \lambda(Y \cap (a_k, b_k)) + \lambda(Y \cap (b_k, c)) \leqslant |b_k - a_k| + \delta_k |c - b_k|$. Consequently, $\varepsilon \leqslant \rho(a_k, c) \leqslant \frac{|b_k - a_k|}{|a_k - c|} + \delta_k$ and hence $|b_k - a_k| \geqslant \frac{\varepsilon}{2}|c - a_k|$ for sufficiently large k. But then

$$\rho(a_k, x_k) \geqslant \frac{\varepsilon}{2}\, \frac{\lambda((a_k, b_k) \cap Y)}{|a_k - b_k|} = \frac{\varepsilon}{2}\rho(a_k, b_k) \geqslant \frac{\varepsilon^2}{2}\, \frac{1}{L^2}$$

[the last inequality follows from (4)].

Thus, in each of the cases a) and b), the densities $\rho(a_k, x_k)$ and $\rho(x_k, a_k')$ are separated from zero. Now using Lemma 3 and the distortion theorem as in case I we get that the upper density of the set Y at the point x is positive. The contradiction proves the basic lemma.

*$\omega(x)$ denotes the limit set of the orbit $\{f^n x\}_{n=0}^{\infty}$.

†By a homterval we mean an interval $\mathcal{I}$, on which all the iterations $f^n|\mathcal{I}$ are injective.

Proof of Theorem. The assertion required follows immediately from the basic lemma since if $I = X_1 \cup X_2$ is a division of the segment into two f-invariant sets, then these sets are symmetric: $\tau X_k \cup X_k = f^{-1}(fX_k) = X_k$.

Proof of Corollary 1. Let Y be a strongly wandering set of positive measure on which all the iterates f^n are injective. Dividing Y into two subsets of positive measure Y_1 and Y_2 we construct a partition of the whole segment into two invariant subsets of positive measure

$$X_1 = \bigcup_{n=-\infty}^{\infty} f^n Y_1,\ X_2 = I \setminus X_1.$$ Contradiction.

In order to prove Corollary 2, we use the topological picture of the dynamics of continuous transformations of a segment [6]. Combining it with Guckenheimer's theorem on the absence of homtervals [7] and the results of [4, 5], we get the following assertion.

THEOREM A. Let $f:I \to I$ be a unimodal transformation with negative Schwarzian; $f:c \to 1 \to 0$. Then there exists a compact set $A \subset I$ such that $\omega(x) = A$ for a.a. $x \in I$.* In addition one of the following conditions holds:

a) A is an attracting cycle or a neutral periodic point;

b) A is contained in a cycle of a transitive segment;

c) A is a solenoid (i.e., a Cantor set on which f is topologically conjugate to a transitive translation with respect to a group).

Proof of Corollary 2. It is clear that the support of any absolutely continuous invariant measure is contained in A. Consequently, in case a) such a measure does not exist, in case c) if it exists then it necessarily coincides with the unique invariant measure of the transformation $f|A$. Finaly, in case b) what is required follows from the theorem proved.

LITERATURE CITED

1. P. Collet and J.-P. Eckmann, Iterated Maps on the Interval as Dynamical Systems, Birkhäuser, Basel, etc. (1980).
2. M. V. Yakobson, "Ergodic theory of one-dimensional maps," Itogi Nauki i Tekhniki. Seriya Sovr. Probl. Mat., VINITI, 2, 204-232 (1985).
3. D. Sullivan, "Quasiconformal homeomorphisms and dynamics. I," Ann. Math., 122, No. 3, 401-418 (1985).
4. A. M. Blokh and M. Yu. Lyubich, "Attractors of transformations of a segment," Funkts. Analiz Prilozhen., 21, No. 2, 70-71 (1987).
5. A. M. Blokh and M. Yu. Lyubich, "Typical behavior of trajectories of transformations of a segment," Teoriya Funktsii, Funkts. Analiz i Ikh Prilozh., 49, 5-16 (1988).
6. A. M. Blokh, "Dynamical systems on one-dimensional branched manifolds. I," Teoriya Funktsii, Funkts. Analiz i Ikh Prilozh., 46, 8-18 (1986).
7. J. Guckenheimer, "Sensitive dependence on initial conditions for one-dimensional maps," Communs. Math. Phys., 70, 133-160 (1979).
8. J. Milnor, "On the concept of attractor," Communs. Math. Phys., 99, 177-195 (1985).

*That is, f has a unique attractor A in the sense of Milnor [8].