

Dispersing Billiards Without Focal Points on Surfaces are Ergodic

A. Krámlı¹, N. Simányi^{2, *}, and D. Szász^{2, *}

¹ Computer and Automation Institute, Hungarian Academy of Sciences, Budapest

² Mathematical Institute, HAS, H-1364 Budapest, POB 127, Hungary

Abstract. Billiards are considered on two-dimensional, smooth, compact Riemannian manifolds with dispersing scatterers. We prove that these billiards are ergodic if only Vetier's conditions for the absence of focal points hold.

0. Introduction

Consider a two-dimensional twice continuously differentiable, compact, closed oriented Riemannian manifold. A simply connected, open subset of the surface will be called a scatterer if its boundary is a twice continuously differentiable curve with strictly positive geodesic curvature from inside. If we have a number of disjoint scatterers, then we call the complement Q of their union a billiard table.

The billiard on Q is a dynamical system corresponding to a motion with unit velocity along geodesics inside Q combined with elastic reflection at the scatterers, i.e. on ∂Q . In particular, if the surface is the torus, then we recover the celebrated Sinai billiard.

In billiards on a surface with scatterers two kinds of behavior can compete: the sufficiently good mixing one caused by the scatterers and a possibly integrable one inside the surface. In 1982 Vetier (Vi (1982), $i = 1, 2$) was able to give conditions under which no focal points arise and thus mixing prevails. He also established the hyperbolic theory for these billiards. Under his conditions the Lyapunov exponents are uniformly bounded away from zero implying the a.e. existence of fibers and properties called the absolute continuity of the foliations. His main conclusion is that the ergodic components of these billiards are positive V(1987).

Here we prove that under Vetier's conditions the billiard is ergodic. By the traditional Hopf-Sinai strategy this follows from a version of the fundamental theorem and, in fact, this is the main result (Theorem 5.1) of our paper.

The proof of the fundamental theorem we separate into two parts. The chief aim of the first, geometric part is to formulate lemmas permitting us to think and

* Partially supported by the Hungarian National Foundation for Scientific Research, grant No. 819/1 and by Central Research Fund, grant No. 501/5/4/1984

measure in euclidean terms while the second, combinatorial-geometric one should show for any irregular configuration of singularity lines arising at the powers of the Poincaré map of the billiard that their bad effect prohibiting the existence of sufficiently long fibers is locally arbitrarily small.

The first part of the paper consisting of four sections is devoted to the geometric aspects. To make the paper possibly self-contained the first three-sections also recollect definitions and properties from the papers of Vetier. Section 1 summarizes the basic notions and conditions. In Sect. 2 the uniform hyperbolicity and the existence of fibers is discussed. Section 3 introduces and analyses the Poincaré map. In Sect. 4 we deduce the main geometric lemmas to be used in the second part consisting of Sects. 5 and 6. In the previous one, we formulate and prove the fundamental theorem. The proof is closely related to that of Bunimovich and Sinai (1973) and the induction is borrowed from Sinai (1980). Our argument, however – with the lemmas provided – is almost purely combinatorial and its structure is simple: there are only two kinds of discarded sets, all the estimates are relative ones, and we do not have to separate the discussion of neighborhoods of double points. Consequently, the relation of the numerous constants also simplifies and is treated explicitly. Finally, Sect. 6 is devoted to the proof of the ergodicity. Section 7 contains some brief remarks while, for completeness, the Appendix gives the necessary information about measurable fibrations.

An efficient way for the first acquaintance with the essence of the paper is to read Sect. 5 having understood the statements of Lemmas 4.2, 4.6, 4.10, and 4.11.

1. Preliminaries – The Definition of the System “Dispersing Billiards on a Surface” and its Elementary Properties

In the course of this paper we make an attempt to use a system of notations compatible with that of Bunimovich and Sinai’s work [Bu-Si (1973)] on dispersing billiards and with that of Vetier’s two works [V1 (1982)] and [V2 (1982)] on dispersing billiards on a surface. The elementary facts concerning the object studied here will be quoted from the papers cited above. The *boundary of the billiard-table*, which will be denoted by ∂Q , consists of the union of the boundaries of the scatterers.

The state of the particle is determined by its position and velocity. So a *state* x means its position q and a direction v together: $x=(q, v)$. The state x can be identified with a unit vector of the tangent plane at the point q . The set of all states is called the *phase-space* and it is denoted by M .

The projection $\pi: M \rightarrow Q$ is defined in the following way: if $x=(q, v) \in M$, then $\pi(x) = q \in Q$.

In the moment when the particle hits the boundary of the billiard-table, the position q of a particle is well-defined, but its velocity is not. If the particle arrives at the boundary with an incoming velocity v_- , then its outgoing vector v_+ after the reflection in the same moment has the following connection with v_- : the sum of the vectors v_- and v_+ is a tangent vector of the boundary of the billiard-table.

The Riemannian metric of the surface induces an area-measure μ_Q on the billiard-table. The set E^q of all unit vectors of the tangent plane at the point q can be

obviously identified with the boundary of the unit circle in this tangent plane. The length-measure on the boundary of the circle corresponds to a measure σ^q on E^q by this identification. Since $M = \bigcup_{q \in Q} E^q$, the measure μ_Q and the measures $\sigma^q (q \in Q)$ together define a measure μ on M :

$$\mu(X) := \int_Q \sigma^q(X \cap E^q) \mu_Q(dq) \quad (X \subset M).$$

The motion of the particle defines the *dynamics* $\{S^t\}_{-\infty < t < \infty}$ on the phase-space: if the particle is in the state x at time 0 then $S^t x$ means its state at time t ($-\infty < t < \infty$). The states $S^t x$ ($-\infty < t < \infty$) together constitute the *trajectory* of the state x . $\{\pi(S^t x) : t \geq 0\}$ is called a ray.

It is known (cf. e.g. [K-S-F (1980)]) that the measure μ is invariant under the dynamics.

Definition 1.1. The dynamical system $(M, \{S^t\}, -\infty < t < \infty, \mu)$ will be called a billiard on a surface, for brevity a billiard.

Now we enumerate the conditions providing that the machinery developed by Vetier works.

Condition 1.2. *There exists a constant $\tau_{\min} > 0$ such that the distance of two scatterers is greater than τ_{\min} .*

Condition 1.3. *The free path is bounded, i.e. there exists a finite value τ_{\max} such that the length of any geodesic on the billiard-table cannot be longer than τ_{\max} .*

Let $k(q)$ denote the geodesic curvature of the boundary of the billiard-table at q ($q \in \partial Q$), and let $k_{\min} := \min_{q \in Q} k(q) > 0$. Let $K(q)$ denote the Gaussian curvature of the billiard-table ($q \in Q$), and let $K_{\max} := \max_{q \in Q} K(q)$.

Condition 1.4. $K_{\max} \leq 0$ or $(K_{\max} > 0, \sqrt{K_{\max}} \cdot \tau_{\max} < \frac{\pi}{2}$ and $\sqrt{K_{\max}} \tau_{\max} < \arctg(k_{\min} \cdot K_{\max}^{-1/2}))$.

2. Uniform Hyperbolicity and Existence of Fibers

In order to shorten the description of our further considerations let us make some conventions:

- (i) all curves in Q and in M are supposed to be smooth;
- (ii) the *curvature* of a curve $g \subset Q$ always means its geodesic curvature;
- (iii) the constants depending only on the billiard-table are called *universal*; in general; the dependence on the billiard table will not be indicated;
- (iv) if $\{S^t y : t > 0, y \in \tilde{g}\}$ is a family of trajectories, $\pi(\{S^t y : t > 0, y \in \tilde{g}\}) \subset Q$ is called a family of rays.

Definition 2.1. The local Lyapunov exponent $\kappa(q)$ of a family of rays at a point $q \in Q$ is the signed curvature of the orthogonal curve to the family passing through q .

The following lemma describes the evolution of $\kappa(x)$ under the action of $\{S^t\}$:

Lemma 2.2. *Let $\{\pi(S^t y): y \in \tilde{g}, t > 0\}$ be a family of rays and set $\kappa(t) := \kappa(\pi(S^t y))$ for a fixed $y \in \tilde{g}$. If t is not a moment of collision then*

$$\kappa'(t) = -K(\pi(S^t y)) - \kappa^2(t).$$

If t is a moment of collision, then

$$\lim_{\tau \rightarrow t+0} \kappa(\tau) - \lim_{\tau \rightarrow t-0} \kappa(\tau) = \frac{2k(q)}{\cos \varphi},$$

where φ is the angle between the outgoing ray and the outer normal of the scatterer ($-\frac{\pi}{2} < \varphi < \frac{\pi}{2}$) and $q \in Q$.

For the proof of this lemma cf. [V1, pp. 1092–1094].

The following lemma explains the name *local Lyapunov exponent* [V1 (1982)].

Lemma 2.3. *Let us choose a ray of a family of rays, and two points of this ray. Let $t_i (i = 1, 2)$ denote the length-parameter value belonging to these points. Let $d_i (i = 1, 2)$ denote the distance of another ray infinitesimally close to the chosen ray at these points, measured along orthogonal trajectories. If $\kappa(t)$ denotes the local Lyapunov exponent of the family of rays along a chosen ray (as in the preceding lemma), then*

$$\lim_{d_1, d_2 \rightarrow 0} \frac{d_2}{d_1} = \exp\left(\int_{t_1}^{t_2} \kappa(t) dt\right).$$

See Lemma 3.5 of [V1 (1982)].

Lemma 2.4. *There exists a universal constant $\omega > 0$ such that, for a family of rays starting from a curve g having curvature greater than k_{\min} , the local Lyapunov exponent is greater than ω at every point $\pi(S^t y), y \in \tilde{g}$.*

For the proof cf. Theorem 3.1 of [V1 (1982)].

A family of rays starting from a point q (a “degenerate curve”) obviously satisfies the condition of Lemma 2.4 (see Fig. 1).

Lemma 2.5. *There exists a universal constant ω_{\max} such that the local Lyapunov exponent $\kappa(t)$ of a family of rays starting from a point $q \in \partial Q$ is bounded by ω_{\max} at every point of the family of rays just before a collision.*

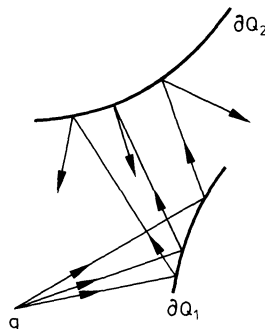


Fig. 1

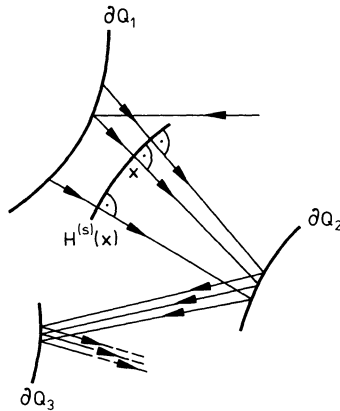


Fig. 2

For the proof cf. Theorem 3.3 of [V1 (1982)].

A consequence of Lemmas 1.1 and 1.2 is the uniform hyperbolicity of a billiard on a surface under the Conditions 1.2–1.4 (cf. Corollary 3.6 of [V1 (1982)]). More precisely the following statement holds:

Lemma 2.6. *If a family of rays starts from a curve \tilde{g} of arc-length l_0 satisfying the condition of Lemma 1.2, then the arc-length of $S^t\tilde{g}$ is bounded from below by $l_0 \cdot \exp(\omega t)$, where ω is the same constant as in Lemma 1.2.*

An analogous statement holds for families of rays starting from a curve g of negative curvature less than $-k_{\min}$ if t is changed by $-t$.

Next we turn to the definition of stable (unstable) fibers. Let g be a curve on the billiard table Q and let $v(q)$ be a unit vector perpendicular to g at each point q ; the curve in the phase space corresponding to this construction will be denoted by \tilde{g} . Obviously $\pi(\tilde{g}) = g$.

Definition 2.7. A curve $H^{(s)}(x)$ ($H^{(u)}(x)$) in M is called a stable (unstable) fiber iff $H^{(s)}(H^{(u)})$ passes through x and for each $y \in H^{(s)}(x)$ ($y \in H^{(u)}(x)$) $S^t y$ ($S^{-t} y$) tends to $S^t x$ ($S^{-t} x$) as $t \rightarrow \infty$.

In [V1 (1982)] the following theorem was proved:

Existence Theorem. *Under the Conditions 1.1–1.3 for almost all $x \in M$ there exist stable and unstable fibers being continuously differentiable curves in M (see Fig. 2)*

Definition 2.8. The set $\bigcup_t \{S^t H^{(s)}(x)\}$ ($\bigcup_t S^t H^{(u)}(x)$) is called a stable (unstable) leaf containing x .

Corollary 2.9. *The arc-length of a stable (unstable) fiber tends to zero exponentially when $t \rightarrow \infty$ ($t \rightarrow -\infty$).*

3. The Poincaré Mapping

The usual way to prove the ergodicity of a flow $\{S^t\}$ is the reduction of the problem to the proof of ergodicity of a discrete time parameter dynamical system – a mapping T – derived from $\{S^t\}$.

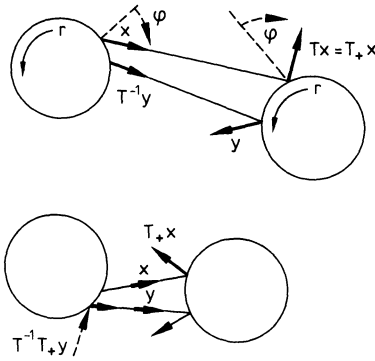


Fig. 3

Definition 3.1. Let ∂M denote the set of such states $y = (q, v)$ that $q \in \partial Q$ and v is directed inwards the billiard table and let, moreover, $\tau(x) = \inf\{t \geq 0 : S^t x \in \partial M\}$ for $x \in M$. Define the mapping $T_+ : M \rightarrow \partial M$ as follows $T_+ x := S^{\tau(x)+0} x$. The restriction T of the mapping T_+ to ∂M is the Poincaré mapping derived from the dynamics $\{S^t\}$. We say that a curve in M is outgoing if $\tilde{g} = T_+ \tilde{g}$.

The dual notion of an outgoing curve is an incoming one which is outgoing in the time reversal dynamics (i.e. $\tilde{g} = T_- \tilde{g}$, where $T_- x := S^{\tau(x)-0} x$).

It is clear that T_+ is invertible (see Fig. 3).

Each component ∂M_i is topologically a cylinder; let us introduce on it the coordinates (r_i, φ_i) , where r is the length parameter of the boundary of the i^{th} scatterer and $-\frac{\pi}{2} \leq \varphi \leq \frac{\pi}{2}$ is the angle of v measured from the outer normal vector of the scatterer (see Figs. 3 and 4).

Lemma 3.2. *The mapping T preserves the measure $dv := \text{const} dr d\varphi \cos \varphi$.*

For the proof see Sect. 4 of [V1 (1982)].

T is obviously continuous on $\partial M \setminus \bigcup_i (S_{i,0} \cup T^{-1} S_{i,0})$, where

$$S_{i,0} = \left\{ \varphi_i = \frac{-\pi}{2} \right\} \cup \left\{ \varphi_i = \frac{\pi}{2} \right\}.$$

Next we summarize the elementary facts concerning the connection between the dynamics $\{S^t\}$ and the mapping T . Their proofs can be found in Sinai's classical paper [S (1970)] and in the majority of later works on billiard systems.

Lemma 3.3. *If $\tilde{g} = \varphi(r) \subset \partial M$ is an outgoing curve and $\kappa_+(q)$ is the curvature of $\pi(\tilde{g})$ in the point $q = \pi(r, \varphi)$, then $\varphi'(r) = \kappa_+(q) \cdot \cos \varphi - k(q)$.*

The following lemma is a consequence of the preceding two ones:

Lemma 3.4. *If \tilde{g} is an incoming (outgoing) curve and $\pi(\tilde{g})$ has a positive (negative) curvature, then $T_+ \tilde{g} (T_- \tilde{g})$ is an increasing (decreasing) curve $\varphi(r)$ in ∂M .*

As a consequence of Lemmas 2.4, 3.3, 3.4, one gets the following two lemmas

Lemma 3.5. *If a curve $g = \pi(\tilde{g})$ in Q has curvature greater (less) than $k_{\min} (-k_{\min})$ then for every $n > 0 (n < 0)$ the curves $\varphi_n(r) := T^n T_+ \tilde{g}$ in ∂M are increasing (decreasing).*

Lemma 3.6. *There exist two universal constants $0 < C_{\text{der}} < C^{\text{der}} < \infty$ such that if an increasing (decreasing) curve $\varphi(r)$ is the T_+ image a curve \tilde{g} satisfying the condition of Lemma 3.5, then*

$$C_{\text{der}} < \frac{d\varphi_n}{dr} < C^{\text{der}} \left(C_{\text{der}} < -\frac{d\varphi_n}{dr} < C^{\text{der}} \right)$$

for $n > 0$ ($n < 0$), where $\varphi_n(r)$ is the same as in Lemma 3.5.

Corollary 3.7. *The curves $TS_0, T^2S_0, \dots, (T^{-1}S_0, T^{-2}S_0, \dots)$ are increasing (decreasing) curves with derivatives between two universal positive (negative) constants.*

Lemma 3.8. *Let us denote by dr_0 and dr'_0 the projection on the r axis of an infinitesimal segment of the curves $\varphi(r)$ and $T\varphi(r)$ at the points (r_0, φ_0) and $(r'_0, \varphi'_0) := T(r_0, \varphi_0)$, respectively. There exists a universal positive constant C_{dil} such that, if $\varphi(r)$ satisfies the condition of Lemma 3.6 and it is increasing, then*

$$\frac{dr'_0 \cos \varphi'_0}{dr_0 \cos \varphi_0} \geq 1 + \frac{C_{\text{dil}}}{\cos \varphi_0}.$$

The statement of Lemma 3.8 is an elementary geometrical fact from the theory of the classical Sinai billiard, where C_{dil} can be chosen $C_{\text{dil}} = \tau_{\text{min}} \cdot k_{\text{min}}$. The proof exploits the local geometry of the collisions, only. Condition 1.4 formulates a global relationship between the curvature of scatterers ∂Q and the Riemannian metric on Q , providing the hyperbolicity of $\{S^t\}$. This condition – roughly speaking – hinders the focusing of a family of rays until the next collision, if it starts from a curve having curvature greater than k_{min} . So analyzing Vetier’s proof of Lemma 2.4 one can check the validity of Lemma 3.8 – with a possibly less but positive constant C_{dil} – for the billiards on a bent able, too.

Corollary 3.9. *Let $dr_0^{(n)}$ denote the projection on the r axis of an infinitesimal segment of the curve $T^n\varphi(r)$ at the point $(r_0^{(n)}, \varphi_0^{(n)}) := T^n(r_0, \varphi_0)$. Then*

$$\frac{dr_0^{(n)} \cos \varphi_0^{(n)}}{dr_0} > C_{\text{dil}}(1 + C_{\text{dil}})^{n-1}.$$

4. Main Geometric Lemmas

First we notice that the ergodicity of the dynamical system $(M, \{S^t\}, \mu)$ can be deduced from that of $(\partial M, T, \nu)$. Moreover, for this end it is enough to prove the ergodicity of the N^{th} iterate T^N , where N is an arbitrary natural number. For a technical reason we choose N in such a way that $C_{\text{dil}}(1 + C_{\text{dil}})^N > 9$.

From the point of view of the geometrical structure there is only one difference between T and T^N , namely: the set of discontinuity lines of T^N , $S_N := \bigcup_{k=0}^N T^{-k}S_0$, has points in which four segments of S_N intersect each other. This fact is the consequence of the possible double tangencies of the rays (see Fig. 4). We assume that there are no tangencies with multiplicity more than 2, which is true for a typical billiard table.

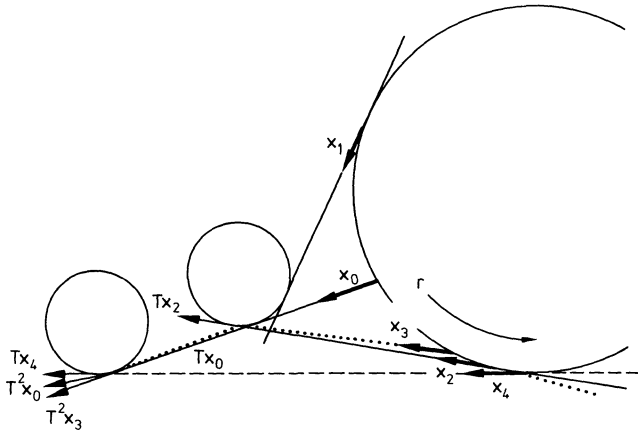
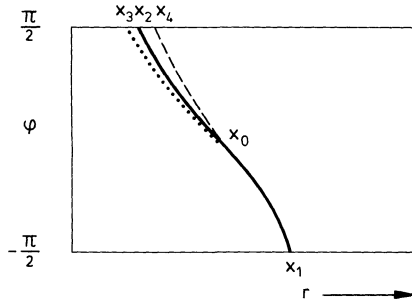


Fig. 4

From now on we will only consider T^N and its powers and for simplicity we use T instead of T^N and e.g. S_k instead of S_{kN} . Moreover, the choice of N provides that for every increasing curve $\varphi(r)$ satisfying the conditions of Lemma 3.8,

$$\frac{dr' \cos \varphi'}{dr} > 9 \quad \text{and} \quad \frac{dr' \cos \varphi'}{dr \cos \varphi} > 9,$$

where (r', φ') already refers to the new mapping T .

Let us introduce two types of measures for the length of increasing (decreasing) curves $\varphi(r)$ having derivatives between two positive (negative) constants.

Lemma 3.6 guarantees that all curves occurring in our arguments will satisfy this condition with constants $\pm C_{\text{der}}$ and $\pm C^{\text{der}}$. So in the sequel – speaking about monotonous curves – we shall not mention this condition.

Definition 4.1. Let $\gamma := \varphi(r)$ be a monotonous curve in ∂M ; denote by $\varrho(\gamma)$ the length of the projection of γ on the r axis and set $\hat{\varrho}(\gamma) := \int_{(r, \varphi) \in \gamma} \cos \varphi dr$.

Let us formulate a lemma on the structure of singularity lines.

Lemma 4.2. *There exists a universal positive constant ε_0 such that for every increasing curve γ and for every quadruple of points $x_1, x_2, x_3, x_4 \in \gamma \cap S_1$ ordered by the coordinate φ ,*

$$\max(\varrho(x_1, x_2), \varrho(x_2, x_3), \varrho(x_3, x_4)) > \varepsilon_0.$$

The following lemma justifies the introduction of the notion of $\hat{\varrho}$ -distance.

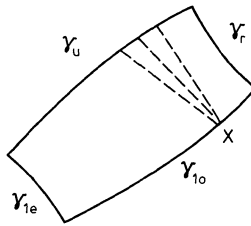


Fig. 5

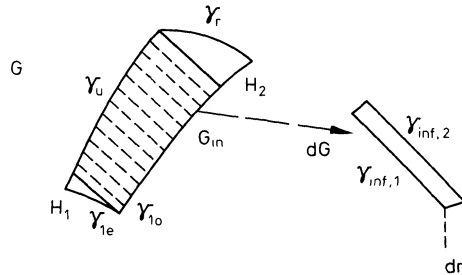


Fig. 6

Lemma 4.3. *If $d\tilde{g}$ is an infinitesimal piece of an incoming curve, then the arc-length of $\pi(d\tilde{g}) = \hat{\varrho}(T_+ d\tilde{g})$.*

Definition 4.4. The $\hat{\varrho}$ -distance between a point $x \in \partial M$ and a monotonous curve γ is defined by the formula

$$\hat{\varrho}(x, \gamma) := \min_{\{\tilde{\gamma}: \tilde{\gamma} \text{ has the opposite monotonicity as } \gamma \text{ and connects } x \text{ with } \gamma\}} \hat{\varrho}(\tilde{\gamma}).$$

Definition 4.5. A quadrilateral G in ∂M with increasing lower and upper sides and decreasing left and right sides (see Fig. 5) is said to be C -regular if

$$\frac{\sup_{x \in \gamma_{1o}} \hat{\varrho}(x, \gamma_u)}{\inf_{x \in \gamma_{1o}} \hat{\varrho}(x, \gamma_u)} < C.$$

G is strongly C -regular if, moreover,

$$\frac{\sup_{\gamma \in \gamma_{1e}} \{\varrho(\gamma, \gamma_r)\}}{\inf_{\gamma \in \gamma_{1e}} \{\varrho(\gamma, \gamma_r)\}} < C.$$

Lemma 4.6. *For every $C > 1$ there exists a number $C_{\text{est}} > 1$ depending only on C such that for any strongly C -regular quadrilateral G ,*

$$C_{\text{est}}^{-1} \varrho(\gamma_{1o}(\mathcal{G})) \cdot \sup_{x \in \gamma_{1o}(\mathcal{G})} \{\hat{\varrho}(x, \gamma_u(\mathcal{G}))\} \leq v(\mathcal{G}) \leq C_{\text{est}} \varrho(\gamma_{1o}(\mathcal{G})) \cdot \sup_{x \in \gamma_{1o}(\mathcal{G})} \{\hat{\varrho}(x, \gamma_u(\mathcal{G}))\}.$$

Proof. An easy Riemannian integral approximation works. The regularity of \mathcal{G} enables us to construct an arbitrarily fine grid in \mathcal{G} , using only piecewise linear increasing and decreasing curves which connect the neighboring grid points to each other by straight lines (see Fig. 6). We can also suppose that each cell of the grid is $(1 + \eta)$ -regular with arbitrary small positive η . Thus, by the definition of the

measure ν for every cell G' we have

$$C'_{\text{est}}{}^{-1} \varrho(\gamma_{10}(\mathcal{G}')) \cdot \hat{\varrho}(\gamma_{1e}(G')) \leq \nu(\mathcal{G}') \leq C'_{\text{est}} \varrho(\gamma_{10}(\mathcal{G}')) \cdot \hat{\varrho}(\gamma_{1e}(\mathcal{G}')).$$

Here the constant C'_{est} is universal. Summing up this inequality for all cells constituting the grid and using the strong C -regularity of G we can find an appropriate C_{est} for which the statement of the lemma holds.

Remark. The upper estimation can be derived supposing the C -regularity, only.

The image of an increasing (decreasing) curve γ under the mapping $T^{-1}(T)$ is not necessarily increasing (decreasing).

Definition 4.7. A curve γ is n -decreasing (n -increasing) if $T\gamma, T^2\gamma, \dots, T^n\gamma(T^{-1}\gamma, T^{-2}\gamma, \dots, T^{-n}\gamma)$ are decreasing (increasing) curves, too.

Lemma 4.8. Let $U_x \subset \partial M$ be a neighborhood of x . If T^n is continuous on U_x , then there exists an n -decreasing curve γ_x passing through x , such that $\gamma_x \cap$ boundary of $U_x \neq \emptyset$.

Proof. Take, in $T^n U_x$, the vertical line γ^* defined by the equation $r = r(T^n x)$ up to the boundary of $T^n U_x$. Then $\gamma_x = T^{-n} \gamma^*$ fulfills the desired conditions in the lemma.

Definition 4.9. A quadrilateral G with increasing lower and upper sides γ_{10} and γ_u and with 1-decreasing left and right sides γ_{1e} and γ_r , is called n -narrow of δ -size iff $\max(\varrho(\gamma_{10}), \varrho(\gamma_u)) > 2^{-n} \delta$ and, for every $x \in \gamma_{10}$ the inequality $\hat{\varrho}(x, \gamma_u) < 8^{-n} \delta$ holds.

Lemma 4.10 (Cutting). Let G be an n -narrow, C -regular quadrilateral of δ -size such that T is continuous on G . Then there exists a constant C_{cut} depending on C such that, if $G_i \subset T(G)$ is any quadrilateral having power (upper) side $\gamma_{i,10}$ ($\gamma_{i,u}$) greater than $2^{-n-1} \cdot \delta$ which is a subset of the lower (upper) side of $T(G)$, then there exist two quadrilaterals $G_{1,i}$ and $G_{2,i}$ such that $G_i \setminus (G_{1,i} \cup G_{2,i})$ has 1-decreasing left and right sides, and, moreover,

$$\frac{\nu(G_{1,i} \cup G_{2,i})}{\nu(G_i)} < \frac{C_{\text{cut}}}{4^n}.$$

Proof. Let us assume that the left side of G has $\hat{\varrho}$ -length $\delta' \cdot 8^{-n}$ ($\delta' < \delta$), and the lower side of G_i has ϱ -length $\delta'' \cdot 2^{-n}$ ($\delta'' > \delta$). It follows from Lemma 3.8, the choice of N and the C -regularity of G that the ϱ -lengths of the left and right sides of G_i are less than $C \cdot 8^{-n} \cdot \delta'$ (see Fig. 7).

There exists a universal constant C_u such that in order to obtain a quadrilateral $G_i \setminus (G_{1,i} \cup G_{2,i})$ having 1-decreasing left and right sides it is enough to choose $G_{1,i}$ and $G_{2,i}$ such that all their sides are less than $C_u \cdot C_{\text{reg}} \cdot 8^{-n} \delta'$. Applying Lemma 4.3 we get the estimate:

$$\frac{\nu(G_{1,i} \cup G_{2,i})}{\nu(G_i)} < \frac{C_u C_{\text{est}} \cdot C \cdot 8^{-n} \delta'}{2^{n-1} \cdot \delta} < C_{\text{cut}} \cdot 4^{-n}.$$

Lemma (Shape) 4.11. For every C_0 there exists a constant C_1 such that for every n if G is a C_0 -regular quadrilateral with increasing lower and upper sides and

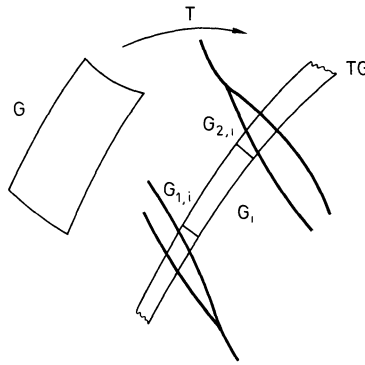


Fig. 7

n-decreasing left and right sides such that T^n is continuous on G , then $T^n G$ is C_1 -regular.

Proof. We shall follow the ideas of Vetier’s proof of the absolute continuity of the fibers for $\{S^t\}$. To do this we should quote the necessary information from [V2 (1982)].

Let $d\tilde{g}$ be an infinitesimal curve in M , passing through the state $x \in M$. The pair $a := (x, l)$ is called the state-curvature pair defined by $d\tilde{g}$, if the curvature of $d\tilde{g}$ is equal to $l, |l| > k_{\min}$. Let $a(t)$ be the state-curvature pair defined by $S^t d\tilde{g}$ ($t > 0$). Set $I(a, t_x) = \int_0^{t_x} a(s) ds$. Lemma 1 of [V2 (1982)] states:

If $a_i = (x_i, l_i)$ is a state-curvature pair ($i = 1, 2$) and the states $S^s x_1$ and x_2 belong to the same stable leaf, then the limit

$$\Delta I(a_1, a_2) := \lim_{t \rightarrow \infty} (I(a_1, t + s) - I(a_2, t))$$

exists.

The proof of this statement is based on the uniform hyperbolicity of $\{S^t\}$. Consequently, for $|s| < \tau_{\max}$, $\Delta I(a_1, a_2)$ is bounded by an absolute constant C_A . Due to Lemma 2.6,

$$C_A^{-1} < \frac{\text{arc-length } \pi(S^t d\tilde{g}_1)}{\text{arc-length } \pi(d\tilde{g}_1)} \cdot \frac{\text{arc-length } \pi(S^t d\tilde{g}_2)}{\text{arc-length } \pi(d\tilde{g}_2)} < C_A.$$

Notice that analogous assertions are true for state-curvature pairs with positive ($l > k_{\min}$) curvature lying on the same unstable leaf if $t \rightarrow -\infty$. In order to prove the lemma we need these “reversed” statements.

Let x_1 and x_2 be two points in $T^n G$. Suppose that γ_1 and γ_2 are two 1-decreasing curves joining the lower and upper sides of G (see Fig. 8) passing through x_1 and x_2 , respectively (see Fig. 8).

Consider two infinitesimal curves $dg_1 = \pi(d\tilde{g}_1)$ and $dg_2 = \pi(d\tilde{g}_2)$ with curvatures less than $-k_{\min}$ such that $x_i \in T^{-1} T_+ g_i \subset \gamma_i$ ($i = 1, 2$). So the above reasoning could be applied to a_i defined by x_i and \tilde{g}_i if x_1 and x_2 were on the same unstable leaf. Vetier’s proof exploits in fact only the exponential convergence of the trajectories belonging to the same leaf. Therefore in order to derive Lemma 4.11 from Vetier’s

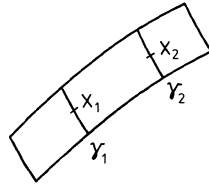


Fig. 8

lemma it is enough to prove the uniform exponential convergence of the trajectories $S^{-t}x_1$ and $S^{-t}x_2$ until the n^{th} collision in the past. But this “finite” convergence is a consequence of the condition that T^n is continuous on G .

Finally, observe that replacing by 1-decreasing curves the decreasing curves connecting x with γ_{10} in the definition of $\hat{\rho}(x, \gamma_{10})$ we make a multiplicative error bounded by an universal constant (remember that the slope of a decreasing curve is between two universal negative constants) κ_1 . So one can choose $C_1 = C_0 \cdot C_A \cdot \kappa_1$.

5. The Fundamental Theorem

In this section we formulate and prove the so-called Fundamental Theorem for dispersing billiards on surfaces. The origin of this theorem for Sinai-billiards can be found in [Sin (1970)] and in [Bu–Sin (1973)].

Theorem 5.1. (Fundamental Theorem for Dispersing Billiards). *For every positive number α ($\alpha < 1$) there exist a positive number $C_\alpha > 1$ and a positive integer k_α such that for every quadrilateral G with left, right, lower and upper sides, γ_{10} , γ_r , γ_{10} , γ_u respectively and for every $\delta > 0$, if*

- (a) γ_u and γ_r are $k_\alpha + 1$ decreasing;
- (b) $C_\alpha^{-1} \delta \leq \min(\rho(\gamma_{10}), \rho(\gamma_u)) \leq \max(\rho(\gamma_{10}), \rho(\gamma_u)) \leq C_\alpha \delta$, and for every $x \in \gamma_{10}$ $C_\alpha^{-1} \delta \leq \hat{\rho}(x, \gamma_u) \leq C_\alpha \delta$;
- (c) $T^{k_\alpha + 1}$ is continuous on G ,

then $\nu(G_\infty) \geq \alpha \nu(G)$, where $G_\infty = \{x \in G : \text{there exists a local stable fiber in } G \text{ passing through } x \text{ and connecting } \gamma_{10} \text{ with } \gamma_u\}$.

Proof. First we notice that there exists a universal constant C_2 such that $\max(\rho(\gamma_{10}), \rho(\gamma_u)) < C_2$; so δ must be less than $C_\alpha \cdot C_2 := \delta_{\max}$. Let $0 < \alpha < 1$ be fixed. Suppose that the numbers $C_\alpha > 1$, k_α and the quadrilateral G are given in such a way that the conditions (a)–(c) are fulfilled. Proving the inequality $\nu(G_\infty) \geq \alpha \nu(G)$ we shall get some necessary conditions on the numbers C_α and k_α which can be satisfied by an appropriate choice of them.

In fact, we shall give an algorithm for a construction of a subset of G . Let us start the algorithm from the quadrilateral $T^{k_\alpha}G$. For every integer $k \geq k_\alpha$ we construct a family $\mathcal{G}^{(k)} := \{G_j^{(k)} : j = 1, 2, \dots, n_k\}$ in such a way that

- (i) for every $k \geq k_\alpha$ the elements of $\mathcal{G}^{(k)}$ are mutually disjoint, C_1 -regular, k -narrow quadrilaterals of δ -size,
- (ii) for every $k \geq k_\alpha$, $G_k := \cup G^{(k)} \subset T^k G$ and $G_k \supset T^{-1} G_{k+1}$,
- (iii) for every $k \geq k_\alpha$ and $1 \leq j \leq n_k$, $\gamma_{10}(G_j^{(k)}) \subset T^k \gamma_{10}(G)$ and $\gamma_u(G_j^{(k)}) \subset T^k \gamma_u(G)$,
- (iv) for every $k \geq k_\alpha$, T is continuous on G_k .

Here the number C_1 is given by Lemma 4.11 using $C_0 = C_\alpha^2$. The construction is going on as follows: Put $k = k_\alpha$, $n_{k_\alpha} = 1$ and $G_1^{(k_\alpha)} = T^{k_\alpha}G$. Now only (i) can be problematic. If C_α and k_α satisfy the condition

$$\left(\frac{9}{8}\right)^k > C_\alpha, \tag{1*}$$

then (i) is fulfilled which can be seen using the condition imposed on N , (the expansion (contraction) coefficient for both the metrics ϱ and $\hat{\varrho}$ is greater (less) than $9 (1/9)$) and using Lemma 4.11.

Suppose that the construction of the systems $\mathcal{G}^{(l)}$ of quadrilaterals is done for $l = k_\alpha, k_\alpha + 1, \dots, k$. We construct $G^{(k+1)}$ as follows. We start from the connected components of the set $TG_k \setminus S_1$ (remember that S_1 is the set of discontinuities of T). Each connected component G' is a polygon whose boundary consists of a finite number of smooth increasing or decreasing curves. Let $G'' \in \mathcal{G}^{(k)}$ be the unique element of $\mathcal{G}^{(k)}$ for which $TG'' \supset G'$. We keep the polygon G' if its boundary contains two connected increasing curves such that the ϱ -length of the lower one is greater than $\delta \cdot 2^{-k}$, otherwise we drop the entire G' . For a remaining G' we apply Lemma 4.10 with the following cast $G = G''$, $n = k$, and $G_i = G'$. So we drop two polygons (usually quadrilaterals) from G' in such a way that G' becomes a C_1 -regular $(k + 1)$ -narrow quadrilateral $G_j^{(k+1)}$ of δ -size. Of course (ii)–(iv) are satisfied. (Lemma 4.11 ensures the C_1 -regularity.) Thus the construction of the systems of quadrilaterals $G^{(k)}$ is complete. The set $P := \bigcap_{k=k_\alpha}^\infty T^{-k}G_n$ is a subset of the set G_∞ , so it is enough to estimate $\nu(G \setminus P)$ from above. This is the purpose of the rest of this section.

There are two types (I and II) of the “loss” of measure:

I. *The dropped polygons (quadrilaterals) from the quadrilaterals G' according to Lemma 4.10.* The total loss of this type is less than

$$\nu(G) \cdot C_{\text{cut}} \sum_{k=k_\alpha}^\infty 4^{-k} = \nu(G) \cdot C_{\text{cut}} 4^{1-k_\alpha} \cdot \frac{1}{3} < \frac{1-\alpha}{2} \nu(G) \tag{A}$$

if

$$4^{k_\alpha} > \frac{8 C_{\text{cut}}}{3(1-\alpha)}. \tag{2*}$$

II. *The entirely dropped quadrilaterals G' .*

In this case $\varrho(\gamma_{10}(G')) \leq \delta \cdot 2^{-k}$ ($\gamma_{10}(G')$ can be empty). From the $(k + 1)$ -narrowness of the quadrilateral G' it follows that $\varrho(\gamma_u(G')) \leq \delta \cdot 2^{-k+1}$. Let us call G'' the father (or ancestor with rank 1) of G' . The father of G'' is called the ancestor with rank 2 of G' etc. We get the *chain of ancestors* with rank 1, 2, ... Let G^* be the unique ancestor of G' with the smallest rank (say l) for which $\varrho(\gamma_{10}(G^*)) > \varepsilon_0$, where ε_0 is the constant in Lemma 4.2. If there is no ancestor G^* of G' for which $\varrho(\gamma_{10}(G^*)) > \varepsilon_0$, then we take $G^* := T^{k_\alpha}G = G_1^{(k_\alpha)}$ (with rank $k + 1 - k_\alpha$). We say that G^* is the forefather of G' . We want to estimate the ratio $\nu(G')/\nu(G^*)$. If $G^* \neq G_1^{(k_\alpha)}$, Lemma 4.6 gives us

$$\frac{\nu(G')}{\nu(G^*)} < C_{\text{est}}^2 \cdot \frac{\delta_{\text{max}}}{\varepsilon_0} \cdot 2^{-k} \cdot 9^{-l}. \tag{1}$$

By virtue of Lemma 4.11, the C_1 -regularity of G^* holds for a well determined C_1 and, if k_α (lower bound for the index $k + 1 - l$ of G) is large enough, then G^* is strongly C_1 -regular as well because it is $(k + 1 - l)$ -narrow. As far as the strong C_1 -regularity of $T^{-l}G'$ is concerned, we see that the usual C_1 -regularity of $T^{-l}G'$ is weaker than that of G^* and the strong version of it is not necessary for the upper bound of the measure $\nu(T^{-l}G')$ in Lemma 4.6. If $G^* = G_1^{(k_\alpha)}$, then Lemma 4.6 gives us

$$\frac{\nu(G')}{\nu(G^*)} < \frac{C_{\text{est}}^2 C_\alpha \cdot 2^{-k} \cdot 9^{-l}}{9^{k_\alpha}}. \tag{2}$$

In this case we can say about the applicability of Lemma 4.6 the same as above. The estimation (1) is weaker than (2) if

$$9^{k_\alpha} > C_\alpha \frac{\varepsilon_0}{\delta_{\text{max}}}. \tag{3^*}$$

So we consider only (1).

We should estimate the sum $\sum_{G'} \frac{\nu(G')}{\nu(G^*)}$ from above for a fixed forefather G^* . This is the kernel of the proof of Theorem 2.1. Let m denote the *index* of the forefather G^* , i.e. $G^* \in \mathcal{G}^{(m)}$. For an estimation of the sum $\sum_{G'} \frac{\nu(G')}{\nu(G^*)}$ we must investigate the father-son relationship between two quadrilaterals G'' and G' . (G'' is the father and G' is his son). The discontinuity lines S_1 split the image TG'' of G'' into many *pieces* (connected components). Some of them are connecting two branches of S_1 and we call them long because the ϱ -length of their lower side is greater than ε_0 (cf. Lemma 4.2). The other pieces are called short. We say that the father-son relationship between G'' and G' is of the first type iff there is no long piece of TG'' with φ coordinates greater than the φ coordinate of G' . Otherwise the relationship is called of the second type. The quadrilateral G' is called of the first type iff every father-son relationship in its chain of ancestors $G^* \rightarrow G'$ is of the first type, otherwise G' is called a quadrilateral of the second type. Figure 9 helps the reader to clarify the different types of father-son relationship.

First we give an upper estimate for the sum

$$\sum_{\substack{G' \text{ is a quadrilateral} \\ \text{of the first type}}} \frac{\nu(G')}{\nu(G^*)}$$

(G^* is a fixed forefather of G'). Let us observe that for a fixed father G'' there are at most four of his sons with father-son relationship of the first type. The reason of this assertion is that at most four branches of S_1 go to a double point of S_1 , moreover, at most three of them go into that point from the left direction (smaller r coordinates) and an analogous statement is true for the right direction. So the number of quadrilaterals G' for which the fixed quadrilateral G^* is their forefather with rank l (i.e. the index of G' is equal to $m + l$) is at most 4^l . Using (1) we get

$$\begin{aligned} \sum_{\substack{G' \text{ is a quadrilateral} \\ \text{of the first type}}} \frac{\nu(G')}{\nu(G^*)} &< \sum_{l=1}^{\infty} \frac{C_{\text{est}}^2 \delta_{\text{max}} \cdot 2^{-m-l+1} 8^{-l^4}}{\varepsilon_0} \\ &< \frac{2 \cdot C_{\text{est}}^2 \delta_{\text{max}}}{3\varepsilon_0} \cdot 2^{-m}. \end{aligned} \tag{3}$$

Let us consider a quadrilateral G' of the second type for which G^* is its forefather. In this case there is a unique quadrilateral G^{**} in the chain of ancestors $G^* \rightarrow G'$ with the largest index such that the father-son relationship between G^{**} and its son is of the second type. Let m_1 denote the index of G^{**} . Among the long pieces of G^{**} there is a unique one (denoted by G^{***}) which is the right neighbour of the son of G^{**} being an element of the chain of ancestors $G^* \rightarrow G'$. (The pieces are ordered by the coordinate r from the left to the right.) We say that G^{***} is the godfather of G' . (Obviously G^{***} is not an element of the chain of ancestors $G^* \rightarrow G'$.) Let the index of the running G' be equal to $m_1 + l$ ($l \geq 1$). The estimation (1) can be used to estimate

the relative measure $\frac{v(G')}{v(G^{***})}$:

$$\frac{v(G')}{v(G^{***})} < \frac{\delta_{\max} \cdot C_{\text{est}}^2 \cdot 2^{-m_1-l+1} \cdot 9^{-l+1}}{\varepsilon_0}. \tag{4}$$

For a fixed G^{***} and $l \in \mathbb{N}$ there are at most 4^l suitable quadrilaterals G' . The reason is again the structure of discontinuity lines S_1 . This fact and (4) imply that for a fixed godfather G^{***} ,

$$\sum_{\substack{G' \\ G^{***} \text{ is a godfather} \\ \text{of } G'}} \frac{v(G')}{v(G^{***})} < \frac{C_{\text{est}}^2 \cdot \delta_{\max} \cdot 2^{-l+1} \cdot 9^{-l+1} \cdot 4^l}{\varepsilon_0} < \frac{C_{\text{est}} \cdot \delta_{\max}}{3\varepsilon_0} \cdot 2^{4-m_1}. \tag{5}$$

Summing up (5) from $m_1 = m$ to $m_1 = \infty$ we get

$$\sum_{\substack{G' \\ G^* \text{ is a quadrilateral} \\ \text{of the second type} \\ \text{whose forefather is } G^*}} \frac{v(G')}{v(G^*)} < \frac{C_{\text{est}}^2 \cdot \delta_{\max}}{\varepsilon_0} \cdot 2^{5-m}. \tag{6}$$

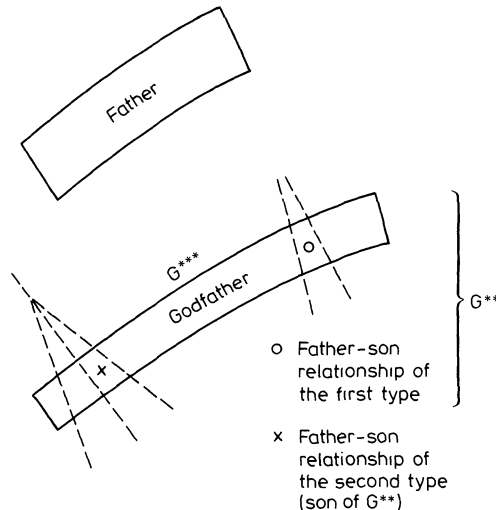


Fig. 9

Inequalities (3) and (6) result in

$$\sum_{\substack{G' \\ G^* \text{ is the forefather} \\ \text{of } G'}} \frac{\nu(G')}{\nu(G^*)} < \frac{34 \cdot C_{\text{est}}^2 \cdot \delta_{\text{max}}}{\varepsilon_0} \cdot 2^{-m}. \tag{7}$$

This shows that the ν -measure of the entirely dropped quadrilaterals G'' divided by $\nu(G)$ is less than

$$\frac{34 \cdot C_{\text{est}}^2 \cdot \delta_{\text{max}}}{\varepsilon_0} \sum_{m=k_\alpha}^{\infty} 2^{-m} = \frac{68 \cdot C_{\text{est}}^2 \cdot \delta_{\text{max}}}{\varepsilon_0} \cdot 2^{-k_\alpha},$$

so if

$$2^{k_\alpha} > \frac{132 \cdot C_{\text{est}}^2 \cdot \delta_{\text{max}}}{3 \cdot (1 - \alpha)\varepsilon_0}, \tag{4*}$$

then the relative loss of ν -measure is again less than $\frac{1 - \alpha}{2}$. Collating this result with (A) we obtain that $\frac{\nu(P)}{\nu(G)} < \alpha$ which we had wanted to prove. The proof is correct if the inequalities (1*)–(4*) are fulfilled, but this can evidently be achieved.

The dual of Theorem 5.1 is, of course, also true. To obtain it we just make the following changes: $\gamma_{\text{le}} \leftrightarrow \gamma_{\text{lo}}$, $\gamma_r \leftrightarrow \gamma_u$, decreasing \leftrightarrow increasing, $T^{k_\alpha+1} \leftrightarrow T^{-(k_\alpha+1)}$, stable \leftrightarrow unstable. Since both forms of the theorem will be applied, we will use the notation $G_\infty^{(s)}$ for G_∞ appearing in the original form and the notation $G_\infty^{(u)}$ arising in the dual one. The constants C_α and k_α can be chosen equal in both versions.

The proof of the ergodicity will be based on the following

Corollary 5.2. *For every α ($0 < \alpha < 1$) there exist a number $C_\alpha > 1$ and a positive integer k_α such that, for every quadrilateral G , for every $\delta > 0$ and for every $n \in \mathbb{N}$, the conditions*

- a) $\gamma_{\text{le}}(G)$ and $\gamma_r(G)$ are $k_\alpha + n + 1$ -decreasing,
 - b) $T^{k_\alpha+n+1}$ is continuous on G ,
 - c) $C_\infty^{-1} \delta \leq \min\{\varrho(\gamma_{\text{lo}}(T^n G)), \varrho(\gamma_u(T^n G))\} \leq \max\{\varrho(\gamma_{\text{lo}}(T^n G)), \varrho(\gamma_u(T^n G))\} \leq C_\alpha \cdot \delta$
- and, for every $x \in \gamma_{\text{lo}}(T^n G)$, $C_\alpha^{-1} \delta \leq \hat{\varrho}(x, \gamma_u(T^n G)) \leq C_\alpha \cdot \delta$
 imply $\nu(G_\infty) \geq \alpha \nu(G)$.

Remark. This corollary allows G to be thin in the unstable direction (and the dual corollary in the stable direction), a fact that will be important in proving the ergodicity.

6. Ergodicity Follows from the Fundamental Theorem

First we prove a local version of the ergodicity: For almost every point $x \in \partial M$ some neighborhood of x belongs to one ergodic component of the dynamics T .

Lemma 6.1. *If $x \in \partial M \setminus \bigcup_{k=-\infty}^{\infty} S_k$, then there exists a neighborhood $U(x)$ of x such that for every subset $R \subset U(x)$ of ν -measure zero there exists another subset $A_R \subset U(x)$ with the following properties:*

- (i) $A_R \cap R = \emptyset$,
- (ii) $\nu(A_R) = \nu(U(x))$,
- (iii) for any pair of points $y, z \in A_R$,

there exist two sequences of local stable and unstable fibers $\gamma_1^s, \gamma_2^s, \dots, \gamma_k^s$ and $\gamma_1^u, \gamma_2^u, \dots, \gamma_k^u$, respectively, such that

- (a) $\left. \begin{array}{l} \gamma_j^s \subset U(x) \\ \gamma_j^u \subset U(x) \end{array} \right\} j=1, 2, \dots, k,$
- (b) $y \in \gamma_1^s$ and $z \in \gamma_k^u$,
- (c) $\emptyset \neq \gamma_j^s \cap \gamma_j^u \subset A_R$ ($j=1, 2, \dots, k$), $\emptyset \neq \gamma_j^u \cap \gamma_{j+1}^s \subset A_R$ ($j=1, 2, \dots, k-1$).

Remark 6.2. A standard Hopf-Hedlund argument shows that $U(x)$ belongs to one ergodic component of the dynamics T .

Proof of Lemma 6.1. Choose an α ($0 < \alpha < 1$). Then Theorem 5.1 gives us two numbers C_α and k_α . Let $U(x)$ be a quadrilateral containing x as an inner point and not intersecting $S_{k_\alpha+n+1} \cup S_{-k_\alpha-n-1}$. (The number n is a large natural number.) We claim that such a neighborhood $U(x)$ satisfies all the conditions (i)–(iii) for every subset $R \subset U(x)$ of ν -measure zero. First we prove that the subset $A'_R = (\text{Int } U(x) \setminus R) \cap F$ would be good if we dropped the conditions $\gamma_j^s \cap \gamma_j^u \subset A'_R$ and $\gamma_j^u \cap \gamma_{j+1}^s \subset A'_R$ from (c). [Here $F \subset \partial M$ is the set of those points $y \in \partial M$ for which there are local stable and unstable fibers containing the point y as an inner point. Of course, $\nu(F) = 1$.] The conditions (i) and (ii) are obviously satisfied. Let y, z be two points in $A'_R = (\text{Int } U(x) \setminus R) \cap F$. Using the properties $U(x) \cap (S_{k_\alpha+n+1} \cup S_{-k_\alpha-n-1}) = \emptyset$, $y, z \in F$ and applying Lemma 4.8 it is easy to see the existence of a sequence of quadrilaterals $G_1^s, G_1^u, \dots, G_k^s, G_k^u$ and of a sequence of positive numbers $\delta_1^s, \delta_1^u, \dots, \delta_k^s, \delta_k^u$ with the following properties:

- (α) $y \in \text{Int } G_1^s \cap (G_1^s)^{(s)}$, $z \in \text{Int } G_k^u \cap (G_k^u)^{(u)}$,
- (β) $\left. \begin{array}{l} G_j^s \subset U(x) \\ G_j^u \subset U(x) \end{array} \right\} j=1, 2, \dots, k,$
- (γ) G_j^s satisfies the conditions (a)–(c) in Corollary 5.2 with the number δ replaced by δ_j^s ($j=1, 2, \dots, k$),
- (δ) G_j^u satisfies the dual conditions (a)–(c) in the dual of Corollary 5.2 with the number δ replaced by δ_j^u ($j=1, 2, \dots, k$),
- (ϵ) $G_j^s \cap G_j^u$ is a quadrilateral such that $\gamma_{1e}(G_j^s \cap G_j^u) \subset \gamma_{1e}(G_j^s)$, $\gamma_r(G_j^s \cap G_j^u) \subset \gamma_r(G_j^s)$, $\gamma_{1o}(G_j^s \cap G_j^u) \subset \gamma_{1o}(G_j^u)$ and $\gamma_u(G_j^s \cap G_j^u) \subset \gamma_u(G_j^u)$.
- (ζ) $G_j^u \cap G_{j+1}^s$ is a quadrilateral such that $\gamma_{1e}(G_j^u \cap G_{j+1}^s) \subset \gamma_{1e}(G_{j+1}^s)$, $\gamma_r(G_j^u \cap G_{j+1}^s) \subset \gamma_r(G_{j+1}^s)$, $\gamma_{1o}(G_j^u \cap G_{j+1}^s) \subset \gamma_{1o}(G_j^u)$ and $\gamma_u(G_j^u \cap G_{j+1}^s) \subset \gamma_u(G_j^u)$.

The choice of n to be large enough (see the Remark after Corollary 5.2) guarantees the narrowness of G_j^s in the unstable direction and the narrowness of G_j^u in the stable direction, so the quadrilaterals $G_1^s, G_1^u, \dots, G_k^s, G_k^u$ can fulfill (ϵ) in this way. Now Corollary 5.2 ensures the validity of (iii) in Lemma 6.1 without the conditions $\gamma_j^s \cap \gamma_j^u \subset A'_R$ and $\gamma_j^u \cap \gamma_{j+1}^s \subset A'_R$. But up to now only the properties $(G_j^s)_{\infty}^{(s)} \neq \emptyset$ and $(G_j^u)_{\infty}^{(u)} \neq \emptyset$ have been used from Theorem 5.1.

The absolute continuity of the partitions of the phase space ∂M into local stable and unstable fibers gives us that $\nu(A'_R) = \nu(U(x))$, where

$$A'_R = \{y \in A'_R : \text{the relative arc-length measure of the set } \gamma^s(y) \cap A'_R \text{ in the curve } \gamma^s(y) \cap U(x) \text{ equals 1 and an analogous statement holds for } \gamma^u(y)\}.$$

Similarly, $\nu(A_R''') = \nu(U(x))$, where

$$A_R''' = \{y \in A_R'' : \text{the relative arc-length measure of the set } \gamma^s(y) \cap A_R'' \text{ in the curve } \gamma^s(y) \cap U(x) \text{ equals 1 and an analogous statement holds for } \gamma^u(y)\},$$

and so on. The intersection $A_R = A_R' \cap A_R'' \cap A_R''' \cap \dots$ satisfies all the properties (i)–(iii) in Lemma 6.1. Only (iii) can be problematic. The construction of the quadrilaterals $G_1^s, G_1^u, \dots, G_k^s, G_k^u$ for any two points $y, z \in A_R$ is going on as earlier. But now, if γ^s is a piece of a local stable fiber connecting $\gamma_{10}(G_j^s)$ with $\gamma_u(G_j^s)$ and if $\mathfrak{X} \subset \gamma^s \cap (G_j^u)_{\infty}^{(u)}$ is a set, measurable by the arc-length measure on γ^s , then this arc-length measure of \mathfrak{X} is positive if and only if the set

$$\{y \in (G_j^u)_{\infty}^{(u)} : \gamma^u(y) \cap \gamma^s \subset \mathfrak{X}\}$$

has positive ν -measure. This is also a consequence of the famous absolute continuity. Of course, the analogous statement (stable-unstable) holds as well. Using these assertions and the definition of the set A_R we get the validity of (iii) in Lemma 6.1.

Remark 6.3. The condition $x \in \partial M \setminus \bigcup_{k=-\infty}^{\infty} S_k$ in Lemma 6.1 is not necessary; it is enough to suppose that $x \notin S_0$ and x is not a double-point, i.e. x belongs to at most one discontinuity line. If, for instance, $x \in S_k$ for a positive integer k , then we can also connect the points y and z with a chain $\gamma_1^s, \gamma_1^u, \dots, \gamma_k^s, \gamma_k^u$, perhaps crossing the line of S_k by some γ_j^u .

Consequently, if $x \notin S_0$ is not a double point, then some neighborhood of it belongs to one ergodic component. The set of these points is arcwise connected in each component of ∂M . (Topology: The complement of any countable set in a two dimensional connected manifold is arcwise connected.) Consequently, cylinders are subsets of the ergodic components. Now, if there were at least two nonempty invariant sets consisting of cylinders, then a perturbation argument around the joint tangent of scatterers belonging to different ergodic components implies that these two scatterers are in the same component. Hence the ergodicity follows.

7. Remarks

1. Among the geometric lemmas it is only the shape lemma which uses the absolute continuity of the fibration. Analogously as was done in Gallavotti’s lectures G (1975), we could have included in the fundamental theorem an assertion about absolute continuity with a bound on the Radon-Nykodim derivative and could have proven this within the fundamental theorem. Here and in the proof of ergodicity we preferred to shorten the exposition and accepted the absolute continuity as proven by Vetier.

2. Dispersing billiards on a surface are generalizations of Sinai billiards in a smooth potential field. The connection is explained in [V1 (1982), Sect. 6].

3. Vetier’s results are strongly related to the approach of [K–S(1986)]: they generalize Pesin’s theory (cf. [P (1977)]) for a wide class of smooth maps with

singularities and they show that, if the Lyapunov exponents are not zero, then the ergodic components are positive. It is an intriguing question to find conditions when the fundamental theorem extends to their class.

4. In Sects. 1–6 all prerequisites for proving the Bernoulli property has been prepared (cf. [G–O (1974)]); consequently Vetier’s billiard is a Bernoulli system and a fortiori it is a K-system.

Acknowledgements. The authors express their sincere gratitude to Ya. G. Sinai for suggesting the problems and to L. A. Bunimovich and again to Ya. G. Sinai for continued stimulating consultations on billiards during the last decade.

References

- [Bu–Si (1973)] Bunimovich, L.A., Sinai, Ya.G.: On the fundamental theorem in the theory of dispersing billiards. *Matem. Sbornik* **90**, 407–423 (1973)
- [G (1975)] Gallavotti, G.: Lectures on the billiard. *Lecture Notes in Physics*, Vol. 38. Moser, J. (ed.) pp. 236–295. Berlin, Heidelberg, New York: Springer
- [G–O (1974)] Gallavotti, G., Ornstein, D.: Billiards and Bernoulli Schemes. *Commun. Math. Phys.* **38**, 83–101 (1974)
- [K–S (1986)] Katok, A., Strelcyn, J.-M. (with the collaboration of Ledrappier, F. and Przytycki, F.): Invariant manifolds, entropy and billiards; smooth maps with singularities. *Lecture Notes in Mathematics*, Vol. 1222. (Berlin, Heidelberg, New York: Springer 1986)
- [P (1977)] Pesin, A.Ya.: Characteristic exponents of Lyapunov and smooth ergodic theory. *Russ. Math. Surv.* **32**, 55–112 (1977)
- [S (1970)] Sinai, Ya.G.: Dynamical systems with elastic reflections. *Russ. Math. Surv.* **25**, 137–189 (1970)
- [S (1979)] Sinai, Ya.G.: Ergodic properties of the Lorentz gas (in Russian). *Funk. Anal. i Prilož.* **13**, 46–59 (1979)
- [V1 (1982)] Vetier, A.: Sinai billiard in potential field (construction of fibers). Révész, P. (ed.): *Coll. Math. Soc. J. Bolyai* **36**, 1079–1146 (1982)
- [V2 (1982)] Vetier, A.: Sinai billiard in potential field (absolute continuity). Proc. 3rd Pann. Symp. ed. J. Mogyoródi, I. Vincze, W. Wertz. pp. 341–351 (1982)

Communicated by A. Jaffe

Received November 5, 1987

