Ergodic properties of certain systems of two-dimensional discs and three-dimensional balls

1987 Russ. Math. Surv. 42 181

(http://iopscience.iop.org/0036-0279/42/3/R06)

Uspekhi Mat. Nauk 42:3 (1987), 153–174 Russian Math. Surveys 42:3 (1987), 181–207

Ergodic properties of certain systems of two-dimensional discs and three-dimensional balls

Ya.G. Sinai and N.I. Chernov

CONTENTS

§1. Introduction

In this paper we analyse ergodic properties of certain systems of two-dimensional discs and three-dimensional balls that move in the absence of external forces with a constant velocity and that interact by means of elastic collisions. The case of two discs on a torus was fully studied in [17]. The presentation is carried out mainly for a system of n discs on a torus. The necessary changes for a system of three-dimensional balls and other systems are indicated.

Let $(q_1^{(i)}, q_2^{(i)})$ be the coordinates of the centre of the i-th disc, and $(p_1^{(i)}, p_2^{(i)})$ the components of its momentum. The $2n$-dimensional torus $\underbrace{\mathbf{Tor}^2\times\mathbf{Tor}^2\times\ldots\times\mathbf{Tor}^2}_{n}$, from which we remove the union of the interiors of the $n(n-1)/2$ cylinders given by

$$(1)\qquad Q_{ij}\colon\ (q_1^{(i)}-q_1^{(j)})^2+(q_2^{(i)}-q_2^{(j)})^2=(2r)^2 \ (\mathrm{mod}\ 1),$$

$(1\leqslant i<j\leqslant n)$, serves as the configuration space Q of the system of n discs of the same radius r. The cylinder Q_{ij} corresponds to the collision of the i-th and the j-th discs.

The system has the first energy integral $H=\frac{1}{2}\sum (p_j^{(i)})^2$, the value of which is supposed to be fixed: $H=H_0$. Then the phase space is the direct product $\mathfrak{M}=Q\times S$, where S is a $(2n-1)$-dimensional sphere of momenta, given by $H=H_0$. The motion of the discs generates a one-parameter group of transformations $\{T^t\}$, $t\in \mathbf{R}$, of the space $\mathfrak{M}$. The invariant measure μ and the σ-algebra of measurable subsets on $\mathfrak{M}$ are naturally defined.

Since the motion of discs on a torus is being considered, the full momentum $P = (P_1, P_2)$, $P_j = \sum_{i=1}^{n} p_j^{(i)}$ $(j = 1, 2)$ is also a first integral of the system. P defines the velocity of the centre of gravity. For fixed $P \neq 0$ the flow $\{T^t\}$ is the direct product of the conditionally-periodic motion of the centre of gravity and the relative motion of the discs or balls that corresponds to $P = 0$ and to the fixed position of the centre of gravity. With a certain assumption of a purely geometrical character (see §3) we prove the following theorem.

Theorem 1. *Let $P = 0$ and let the position of the centre of gravity be fixed. Then the ergodic components of the flow $\{T^t\}$ have positive measure. The flow is a K-flow on each such component.*

We recall ([10], [22]) that K-flows are ergodic, have mixing of all powers, have positive entropy, and in the orthogonal complement to the one-dimensional subspace of constants the groups of unitary operators adjoint to them have countably-multiple Lebesgue spectrum. As for the geometrical assumption mentioned above, it is verified directly for small values of n $(n \leqslant 10)$. It seems all but certain that it holds for all n, but this has not been proved.

A corollary related to $P \neq 0$ follows from Theorem 1.

Corollary 1. *Let $P \neq 0$ and let the conditionally-periodic flow corresponding to the motion of the centre of gravity be ergodic. Then, under the conditions of Theorem* 1, *the ergodic components of $\{T^t\}$ have positive measure. On each such component the flow is the direct product of the conditionally-periodic flow and the K-flow.*

The following theorems are true without additional assumptions.

Theorem 2. *For a system of n discs (balls) on a torus with $r \leqslant r_n$ there is an open subset Θ such that the ergodic components of $\{T^t\}$ that substantially intersect Θ (that is, the conditional measure of Θ on such a component is positive) have positive measure.*

Theorem 3. *For a system of n discs (balls) on a torus, $\{T^t\}$ has positive entropy.*

A system of n elastically colliding discs or balls in a domain of any form or on a torus is always reducible to a system of billiard type (see [10] and the next section). In the case of domains with a piecewise flat boundary the corresponding billiards is semiscattering, that is, the boundary of the corresponding domain is convex (within), but not strictly convex. Billiards with a strictly convex boundary are called scattering. The properties of scattering billiards are similar to those of hyperbolic systems of the type of Anosov systems, while semiscattering billiards are similar to partially hyperbolic systems (see [7], [22]). The first step in an investigation of the ergodic properties of similar systems consists in the construction of stable and unstable manifolds for "individual" points of the phase space. A suitable technique is a development of the technique of the proof of the Hadamard-Perron theorem, and it has been well developed (see [1], [2], [13], [22]). In our case we initially construct tangent spaces to these manifolds. Such spaces are given by operators of the second fundamental (quadratic) form of the projections of the stable and unstable manifolds on the configuration space. These operators are solutions of the Jacobi equations. In the case of billiard systems the Jacobi equations turn out to be of difference type. Therefore, their solutions are constructed in the form of operator-valued continued fractions. For scattering billiards the operators arising are strictly positive. For semiscattering billiards these operators are merely non-negative. The zero subspaces arising are usually connected with the existence of additional first integrals of motion.

If for almost every point of the phase space the dimension of the positive subspace of the operator of the second fundamental form is maximal, then the local stable and unstable manifolds exist almost everywhere and the corresponding foliations are absolutely continuous. If the sum of the dimensions of the stable and unstable manifolds is one less than the dimension of the whole relevant submanifold of the phase space, it follows by means of reasoning of Hopf type (see [18], [19]) that the ergodic components have positive measure, and on each such component $\{T^t\}$ is a K-flow. It can even be shown by using a finer technique that $\{T^t\}$ is a B-flow, that is, it is metrically isomorphic to the ergodic superstructure (tower) on the Bernoulli automorphism. In precisely this way Theorem 1 and its Corollary are proved.

The further investigation consists in finding out when an ergodic component is unique. In the case of scattering billiards this intention is pursued with the help of "the fundamental theorem of the theory of scattering billiards", first proved in [19]. There are two different proofs of this theorem in [8] and in [20]. In §4 we put forward an additional proof of the "fundamental theorem", which could be applied, with additional assumptions, to certain semiscattering billiards. On the basis of this theorem, in §5 we investigate open domains in the space of parameters of the system of $n = 3$ discs, for which it is shown that the ergodic component is unique. The construction of similar domains for $n > 3$ discs is reduced to the straightforward examination of several degenerate possibilities, which can be done with the help of computers.

It is clear from the above discussion that the results of the present paper are related to the substantiation of the ergodic hypothesis, put forward by Boltzmann more than 100 years ago (see [6]). This hypothesis means in modern terms that a non-linear Hamiltonian system of a general kind on a manifold of constant energy is ergodic. It is certain that Boltzmann related this hypothesis to systems with many degrees of freedom, since in his book [6] he dealt with the ergodic hypothesis in connection with the foundations of statistical mechanics. At that time, of course, there was no notion of thermodynamic limiting transition, when the number of particles grows in proportion to the volume of the system, and the mean distance between the particles is taken to be the unit of length.

The basic notions of the ergodic theory appeared in the mathematical works of Poincaré, Birkhoff, and von Neumann, and the ergodic hypothesis began to be understood as just the hypothesis of ergodicity of this or that dynamical system. The connection with the foundations of statistical mechanics lost its importance, at least temporarily. However, the question of ergodicity became meaningful for finite-dimensional dynamical systems with a small number of degrees of freedom. Individual examples of classical dynamical systems were investigated in this connection, and their ergodicity was proved. One should mention first of all geodesic flows on compact manifolds of negative curvature. One of the first important steps was taken by Hadamard [29], whose results were employed in the related works of Morse [33], Hedlund [30], and Hopf [23]. Geodesic flows on manifolds of negative curvature are one of the principal examples of Anosov systems (see [1], [2]). Ergodicity, mixing, and the K-property of such systems are being investigated at present in sufficient detail (see [1], [2], [18], [22]).

At the beginning of the fifties the famous KAM (Kolmogorov-Arnol'd-Moser) theory of small perturbations of integrable Hamiltonian systems was created (see [3]). One of the principal results of this theory is that under rather general conditions a perturbed Hamiltonian system remains non-ergodic on a set of positive measure. The ergodic theory of Boltzmann in its initial form was thus refuted, but the question arose, which Hamiltonian systems, in particular those of physical origin, are ergodic.

The geodesic flows on compact manifolds of negative curvature discussed above have the typical property of uniform exponential instability of motion, like all Anosov systems. This means, in particular, that geodesic curves that emanate from the same point diverge at an exponential rate. In the forties, the Leningrad physicist Krylov [11] observed that dynamical systems corresponding to the motion of discs or balls with elastic collisions had the same exponential instability. The scattering role of the negative curvature is taken by the convex (within) boundary of the configuration space, composed of cylinders of the type (1). One cannot take the reasoning of Krylov as a rigorous mathematical proof, but he presented the basic idea quite clearly: dynamical systems with elastic collisions should be ergodic because of the same exponential instability as that of geodesic flows in spaces of negative curvature.

The corresponding mathematically rigorous result for a system of two discs or balls was obtained, as we mentioned above, in [19]. The difficulty of investigation is connected with the discontinuous character of the dynamics and non-uniform instability. Precisely these difficulties will be overcome in the next sections of the present paper for the cases described above. The results of the present paper are the first and up to now the only statements on ergodic properties of systems of n discs. Before the appearance of the entropy theory of dynamical systems and the theory of systems with hyperbolic properties of instability there was no approach to similar problems at all. The announcement made in [17] for the general situation must be regarded as premature.

Simultaneously with the investigation of the ergodicity of finite-dimensional systems the development of the mathematical foundations of statistical mechanics was proceeding, in which an analysis of systems of many degrees of freedom, more precisely, an analysis of the properties of such systems under thermodynamic limiting transition, appeared in the foreground. In equilibrium statistical mechanics the main role is played by infinite-dimensional dynamical systems composed of infinitely many identical particles, and by special probability distributions on their phase spaces, which are called limiting Gibbs distributions. These systems first appeared in the paper of Bogolyubov and Khatset [4] (see the modified presentation in [5]), and then their theory was constructed and developed in the works of Dobrushin [9], Lanford and Ruelle [32], Ruelle [16], and others. The basic assumption (the Gibbs postulate) of equilibrium statistical mechanics is that an infinite-dimensional dynamical system of statistical mechanics in a state of thermodynamic equilibrium (that is, in the absence of thermal and dynamical processes) is subject to the limiting Gibbs distribution. At present the problem of substantiating equilibrium statistical mechanics is posed as the problem of explaining the exceptional role of such distributions.

One of the possible approaches to this problem arises in non-equilibrium statistical mechanics. According to the main idea of Bogolubov the evolution of non-equilibrium distributions in systems of statistical mechanics has two clearly distinguished time scales. The first, microscopic, scale is equal to the mean duration of a free run, more precisely, to the time unit of microscopic motions. It is assumed that during such time periods local equilibrium is established in the non-equilibrium system owing to collisions. Local equilibrium means that correlation functions on microscopic distances, that is, on distances of the order of magnitude of the mean distance between the particles, are close to the correlation functions of the limiting Gibbs distribution. However, the parameters of this distribution are not constant but slowly varying functions in space and time. Their evolution occurs on the second, slower, hydrodynamic time scale and is described by equations of hydrodynamic type.

The role of properties of the type of ergodicity and mixing during establishment of local equilibrium is obvious. It is less clear, but nevertheless certain, that the same properties are essential for an investigation of the dynamics of locally-equilibrium distributions. One can hope that the results obtained in the present paper will be helpful in the study of the profound and difficult problems of kinetics of complex systems described above.

We are grateful to S.P. Novikov for useful discussion of the questions studied in the present paper. We also thank L.A. Bunimovich, A. Krámli, D. Sas, and N. Simányi, who carefully read the manuscript and made many important and useful remarks.

§2. Necessary information on scattering and semiscattering billiards

As we mentioned above, a system of n freely moving and elastically colliding discs can be reduced to a billiard system of special type (see [9]). A billiard system or simply billiards is a dynamical system corresponding to the motion of a material point within a bounded domain $Q \subset \mathbf{R}^d$ or $Q \subset \mathrm{Tor}^d$, $d \geqslant 2$. The point moves freely, that is, with a constant velocity, within Q and is reflected from the boundary ∂Q by the law "the angle of incidence is equal to the angle of reflection". The norm of the velocity vector $\|v\|$ corresponding to such a motion is a first integral. The configuration space is $\bar{Q} = Q \cup \partial Q$, and the phase space is $\mathfrak{M} = \bar{Q} \times S$, where S is the $(d-1)$-dimensional sphere of the velocity vectors for which $\|v\| = v_0$. Points of the phase space are denoted by $x = (q, v)$, where $q \in \bar{Q}$, $v \in S$. The natural projection $\mathfrak{M} \to \bar{Q}$ is denoted by $\pi : \pi(q, v) = q$. A billiard system in Q generates a flow $\{T^t\}$ in $\mathfrak{M}$ (see [10]).

The boundary ∂Q is assumed to be piecewise smooth. The latter means that $\partial Q = \partial Q_1 \cup \ldots \cup \partial Q_k$, where the ∂Q_i are smooth submanifolds of codimension 1 that are pairwise mutually transversal at all points of intersection. The flow $\{T^t\}$ is uniquely defined only for points whose trajectories do not pass through the intersections of the smooth boundary components and do not have infinitely many reflections from the boundary during a limited time interval. We denote by $\mathfrak{M}' \subset \mathfrak{M}$ an invariant subset for which $\{T^t\}$ is defined for all t, $-\infty < t < \infty$. Then $\mu(\mathfrak{M} \setminus \mathfrak{M}') = 0$ (see [10]).

The boundary of the phase space $\partial\mathfrak{M} = \partial Q \times S$ plays the main role in our analysis. We put $\mathfrak{M}_1^+ = \{x \in \partial \mathfrak{M}\colon (v, n(q)) \geqslant 0\}$, $\mathfrak{M}_1 = \mathfrak{M}_1^+ \cap \mathfrak{M}'$, where $n(q)$ is the unit vector normal to ∂Q at the point q, directed inside Q. A natural derived automorphism of $\{T^t\}$, $T_1\colon \mathfrak{M}_1 \to \mathfrak{M}_1$, is defined on $\mathfrak{M}_1$. Namely, let $s(x) > 0$ be the first positive time when the semitrajectory $\{T^t x\}$, $t > 0$, reaches the boundary, and $\pi_1\colon \mathfrak{M}' \to \mathfrak{M}_1$ is defined by $\pi_1 x = T^{s(x)+0} x$. Then $T_1 = \pi_1 |_{\mathfrak{M}_1}$.

The invariant Liouville measure μ of $\{T^t\}$ is the direct product $\mu = \mu_Q \times \omega_S$, where μ_Q is Lebesgue measure on Q, while ω_S is Lebesgue measure on S. The invariant measure μ_1 for T_1 has the form $d\mu_1(q, v) = (v, n(q)) dq\, d\omega_S$, where dq is the measure on ∂Q induced by the Riemannian metric.

Ergodic properties of billiard systems depend substantially on geometric properties of ∂Q, more precisely, on its curvature, which is described by an operator of the second fundamental form $K(q)$, $q \in \partial Q$. The operator $K(q)$ is a self-adjoint operator acting in the space $\mathcal{T}_q$ tangent to ∂Q at $q \in \partial Q$.

A billiard system is called scattering (see [19]) if $K(q) > 0$ everywhere on ∂Q. The ergodic properties of scattering billiards have been quite well studied. Namely, they are K-systems (see [19], [20]) and B-systems [28]. In some cases one can investigate the rate of decrease of time correlation functions with the help of Markov partitions (see [25], [26]).

A system of n discs on a torus or in a domain with flat boundaries is reducible to a billiard system for which $K(q) \geqslant 0$, since each cylinder Q_{ij} is flat along a $(2n-2)$-dimensional subspace. Billiards for which $K(q) \geqslant 0$ are called semiscattering, and their analysis is much more complicated. Stable and unstable transversal foliations were constructed for semiscattering billiards under certain assumptions in [24], [31] (see also [23]). These results are described in the next section.

§3. Stable and unstable foliations

We recall that a local stable manifold (LSM) of a point $x \in \mathfrak{M}$ is a C^2-smooth open submanifold $W \subset \mathfrak{M}$ such that:

1) $x \in W$;
2) $\rho(T^t y_1, T^t y_2) \leqslant c_1 \exp\{-c_2 t\} \rho(y_1, y_2)$

for all $y_1, y_2 \in W$ and $t > 0$, where $c_1 = c_1(W) > 0$, $c_2 = c_2(W) > 0$ are constants, and ρ is a Riemannian metric on $\mathfrak{M}$.

A local unstable manifold (LUM) is defined similarly, but now for $t < 0$. The analysis of LSM and LUM is carried out with the help of the corresponding operators of the second fundamental form. Namely, for arbitrary semiscattering billiards in the d-dimensional domain Q we take a C^2-smooth oriented open submanifold $\tilde{\Sigma} \subset Q$ of codimension 1 and its clothing Σ by unit normal vectors. There are two possibilities for such clothing denoted by Σ and $-\Sigma$. We call Σ and $\tilde{\Sigma}$ a local manifold (LM) and its support, respectively. If $\{v(q),\ q \in \tilde{\Sigma}\}$ is the field of unit normal vectors, then an operator $B_\Sigma(x)$, $x = (q, v(q))$, of the second fundamental form is defined by

$$v(q + dq) = v(q) + B_\Sigma(x) dq + o(\| dq \|).$$

Here $B_\Sigma(x)$ is a linear self-adjoint operator, acting in the $(d-1)$-dimensional subspace $J(x)$ tangent to $\tilde{\Sigma}$ at $q \in \tilde{\Sigma}$. We note that $J(x)$ depends only on x, but not on Σ. An LM Σ is called convex (strictly convex) if $B_\Sigma(x) \geqslant 0$ ($B_\Sigma(x) > 0$) for all $x \in \Sigma$. If $B_\Sigma(x) \leqslant 0$ ($B_\Sigma(x) < 0$) for all $x \in \Sigma$, then Σ is called concave (strictly concave).

We now investigate the behaviour of $B_\Sigma(x)$ resulting from the dynamics. We put $x_t = T^t x$, $q_t = \pi(x_t)$ and initially assume that t is so small that $T^s\Sigma = \Sigma_s$ does not intersect $\partial\mathfrak{M}$, $0 \leqslant s \leqslant t$. Then $J(x)$ and $J(x_t)$ are parallel to each other and can be naturally identified. It is easy to check that in this case $B_{\Sigma_t}(x_t) = B_\Sigma(x)(I + tB_\Sigma(x))^{-1}$, where I is a unitary operator (see [17]).

It follows from this formula that $T^t\Sigma$ is convex (strictly convex) if Σ is convex (strictly convex).

Now let t be such that $T^t x \in \partial\mathfrak{M}$ and there is a reflection from the boundary. It makes sense to study vectors v_{t-0}, v_{t+0} before and after the reflection from the boundary, subspaces $J(x_{t-0})$, $J(x_{t+0})$ that are normal to v_{t-0}, v_{t+0} respectively, and operators of the second fundamental form

$$B_{\Sigma_{t-0}}(x_{t-0}),\ B_{\Sigma_{t+0}}(x_{t+0}),\ x_{t\pm 0}=(q_t,\ v_{t\pm 0}).$$

For a point $q_t \in \partial Q$ there is defined the operator of the second fundamental form $K(q_t)$ of ∂Q and of the field of the unit normal vectors $n(q)$. We shall need an operator $U(x_t)$ that maps $J(x_{t+0})$ on $J(x_{t-0})$ parallel to $n(q_t)$, an operator $V(x_t)$ that maps $J(x_{t+0})$ on $\mathcal{T}_0(q_t)$ parallel to v_{t+0}, and an operator $V^*(x_t)$, that maps $\mathcal{T}_0(q_t)$ on $J(x_{t+0})$ parallel to $n(q_t)$, where $\mathcal{T}_0(x_t)$ is the tangent space to ∂Q at q_t. Then (see [34])

$$B_{\Sigma_{t+0}}(x_{t+0})=U^{-1}(x_t)\,B_{\Sigma_{t-0}}(x_{t-0})\,U(x_t)+2\,(v_{t+0},\ n\,(q_t))\,V^*(x_t)\,K\,(q_t)\,V\,(x_t).$$

It follows from the last formula that in the case of semiscattering billiards $B_{\Sigma_t}(x_t) \geqslant 0$ for all $t > 0$ if $B_\Sigma(x) \geqslant 0$, $x \in \Sigma$, that is, the image of a convex LM remains a convex LM under the action of $\{T^t\}$.

For each $x = (q,\ v) \in \mathfrak{M}'$ we define a linear self-adjoint operator $B(x)$, acting in the $(d-1)$-dimensional space $J(x)$. Let $t_0 = 0 < t_1 < t_2 < \dots$ be the instants of successive reflections from the boundary of the semitrajectory $\{T^t x,\ t > 0\}$. We now denote by $K_n = K(q_{t_n})$, $V_n = V(x_{t_n})$, $V_n^* = V^*(x_{t_n})$, $U_n = U(x_{t_n})$, $\cos\varphi_n = (v_{t_n+0},\ n(q_{t_n}))$, $s_n = t_n - t_{n-1}$ the corresponding operators introduced above. We now write an operator-valued continued fraction

$$(2)\qquad B(x)=\cfrac{I}{s_1 I+U_1\cfrac{I}{2\cos\varphi_1 V_1^* K_1 V_1+\cfrac{I}{s_2 I+U_2\cfrac{I}{2\cos\varphi_2 V_2^* K_2 V_2+\dots}U_2^{-1}}}U_1^{-1}}$$

For semiscattering billiards for every $x \in \mathfrak{M}'$ the continuous fraction $B(x)$ exists as the limit of finite continuous fractions. It is a self-adjoint non-negative linear operator in $J(x)$ that depends continuously on $x \in \mathfrak{M}'$ (see [23], [24]). We can thus write a decomposition

$$J(x) = J_+(x) \oplus J_0(x),\ \text{ where }\ B(x)\,|_{J_+(x)} > 0,\ \ B(x)\,|_{J_0(x)} = 0.$$

We put $j(x) = \dim J_+(x)$, $x \in \mathfrak{M}'$. The set

$$\Omega=\{x\in\mathfrak{M}':\ j\,(x)\neq 0\ \text{ and for some neighbourhood } V\,(x)\subset\mathfrak{M}$$
$$\text{the function } j(x) \text{ is constant on } V\,(x)\cap\mathfrak{M}'\}$$

is open in $\mathfrak{M}'$. For every $x = (q, v) \in Q$ we consider the tangent space $\mathcal{T}_x\mathfrak{M} = \mathcal{T}_qQ \oplus \mathcal{T}_vS$, where $\mathcal{T}_vS$ is naturally isomorphic to $J(x)$. The set

$$E(x) = \{(e, f)\colon e \in J_+(x),\ f = -B(x)e\}$$

is a linear subspace of $\mathcal{T}_x\mathfrak{M}$, $\dim E(x) = j(x)$.

***Theorem* 4** (see [24]). *Let Q be such that $K(q) \neq 0$ for some $q \in \partial Q$. Then $\Omega \neq \varnothing$ and for almost every point $x \in \Omega$ there is an LSM $W^{(s)}(x)$, $x \in W^{(s)}(x)$, where $\mathcal{T}_xW^{(s)}(x) = E(x)$.*

An analysis of $j(x)$ is based on the following equality (see [24]):

$$(3)\qquad J_0(x) = \\ = \{w \in J(x)\colon K_lV_lU_l^{-1}U_{l-1}^{-1} \ldots U_1^{-1}w = 0 \text{ for all } l = 1, 2, \ldots\}.$$

Since the phase space is finite-dimensional, there is an $l(x)$ such that

$$(4)\qquad J_0(x) = \\ = \{w \in J(x)\colon K_lV_lU_l^{-1}U_{l-1}^{-1} \ldots U_1^{-1}w = 0 \text{ for all } l = 1, 2, \ldots, l(x)\}.$$

We denote by $l_0(x)$ the minimal permissible $l(x)$. Then $l_0(x)$ is a non-negative integer-valued function on $\mathfrak{M}'$.

If $w \in J_0(x)$, then by (3) for all $m \geqslant 1$

$$V_m^*K_mV_mU_m^{-1}U_{m-1}^{-1} \ldots U_1^{-1}w = 0,$$

that is, $f_m = U_m^{-1} \ldots U_1^{-1}w$ is an eigenvector of the operator $V_m^*K_mV_m$ with eigenvalue 0. This operator is self-adjoint, non-negative, and has a unique eigenvector with a positive eigenvalue: $V_m^*K_mV_me_m = \lambda_me_m$, $\lambda_m > 0$. Therefore $e_m^{(+)} = U_1 \ldots U_me_m \in J_+(x)$ and by (4)

$$(5)\qquad J_+(x) = \mathcal{L}\{e_m^{(+)},\ m = 1, 2, \ldots, l_0(x)\},$$

$\mathcal{L}$ denotes the linear space generated by the corresponding vectors. We have $l_0(x) \geqslant j(x)$.

We investigate $j(x)$ for a system of n discs (balls) on a torus. The boundary ∂Q is the union of cylinders (1). Each cylinder is the direct product of a circle (a sphere in the ball case) and a $(2n-2)$-dimensional linear space ($(3n-3)$-dimensional in the ball case). It is easy to see that the intersection of all these spaces includes a two-dimensional (three-dimensional in the ball case) space generated by the vectors (1, 0, 1, 0, ..., 1, 0) and (0, 1, 0, 1, ..., 0, 1) from $\mathbf{R}^{2n}$. Therefore, by (3) $\dim J_0(x) \geqslant 2$ for all $x \in \mathfrak{M}'$ and thus $j(x) \leqslant 2n-3$ (for balls $j(x) \leqslant 3n-4$). The existence of the general two-dimensional subspace is connected with the preservation of the full momentum P of the system and with the conditionally-periodic motion of the centre of gravity. When $P = 0$ the centre of gravity is stationary, and the corresponding system is also a billiard system $\mathfrak{M}^{(0)}$ in a domain $Q^{(0)}$ of a $(2n-2)$-dimensional torus. The boundary is now composed of the cylinders $Q_{ij}^{(0)} = Q_{ij} \cap Q^{(0)}$. The maximal dimension of LSM's and LUM's can be $2n-3$ ($3n-4$ in the ball case).

Suppose that we have a system of n discs that splits into subsystems of n_1, n_2 discs, $n = n_1 + n_2$, where discs from different subsystems do not interact. According to the above reasoning, for such a situation $j(x) \leqslant \leqslant (2n_1 - 3) + (2n_2 - 3) = 2n - 6$ when $n_1, n_2 \geqslant 2$ and $j(x) = 2(n-1) - 3 = = 2n - 5$ when $n_1 = 1$. In both cases the dimension of an LSM is not maximal.

Let us prove the following statement by induction on n.

***Lemma* 1.** *Let the statement of Corollary* 1 *hold for all* $n' < n$. *Then in a system of* n *discs on a ball the measure of the trajectories along which the discs split into non-interacting groups is equal to zero.*

Proof. Assume that the lemma is false, and that there is a subset $C_0 \subset \mathfrak{M}$ of positive measure that corresponds to the non-interacting groups. We write $x = (x', x'')$, where x', x'' denote points of the phase spaces of the corresponding groups. Let $P' = P(x')$, $P'' = P(x'')$ be the vectors of the full momentum of x', x'' respectively. The components P', P'' are rationally independent of each other for almost every x. We denote by $M'(P')$, $M''(P'')$ the phase spaces of x', x'' with the values of full momentum P', P'' respectively. The flow $\{T^t\}$ in C_0 is the direct product of the flows $\{T^t_{\mathrm{I}}\}$, $\{T^t_{\mathrm{II}}\}$ in $M'(P')$, $M''(P'')$. It follows from the above that for typical P', P'' the ergodic components $\{T^t\}|M'(P')$, $M''(P'')$ have positive measure, and that they are the direct product of spaces of possible positions of the centres of gravity of each of the groups, that is, of two-dimensional tori and subsets of positive measure in spaces of relative positions of each group, when $P' = P'' = 0$. This means, however, that all possible (x', x'') that differ in the position of their centres of gravity occur mod 0 in the same ergodic component. In other words, taking x', x'', we can move them as a whole in arbitrary fashion, remaining in the same ergodic component. But it is clear that there exists a set of shifts of positive measure that is inadmissible, since it leads to a superposition of discs from the different groups. Thus $\mu(C_0) = 0$, and our statement is proved.

The proof of Lemma 1 for the ball case is similar.

Suppose now that it is impossible to split a system of n discs into two non-interacting subsystems for any $t > 0$. Apparently, in this case the condition $j(x) < 2n - 3$ is satisfied only on the union of countably many submanifolds of smaller dimension in the phase space $\mathfrak{M}^{(0)}$ that corresponds to special degenerate trajectories, but we do not have a full proof of this statement for all n. For small values of n ($n \leqslant 10$) it can be shown by explicitly calculating the coordinates of the vectors $e_m^{(+)}$ in (5), selecting $2n - 3$ such vectors ($3n - 4$ vectors in the ball case), and verifying their linear independence for every possible sequence of pairwise collisions of discs. For arbitrary n the necessary statement is reducible to the following:

***Proposition* 1.** *We consider points $x \in \mathfrak{M}^{(0)}$ such that the system of discs cannot be split into non-interacting subsystems when $t > t_0$ for every t_0. Then for any such point that does not belong to the union of some countably many submanifolds of smaller dimension there is an $\tilde{l} = \tilde{l}(x)$ in $\mathfrak{M}^{(0)}$ such that the specification of the velocity vectors of all discs at the instants of the first $\tilde{l}$ collisions and the directions of the lines of centres of colliding discs (centre lines) uniquely determines a point of the phase space.*

We shall prove that $j(x) = 2n-3$ follows from Proposition 1. Suppose this is untrue, that is, $\dim J_0(x) \geqslant 1$. We consider a vector $w_0 \in J_0(x)$, $w_0 \neq 0$, and a point $x' = (q + \varepsilon w_0, v)$, where $(q, v) = x$, while ε is chosen to be so small that the first $\tilde{l}$ reflections of the trajectory of x' are from the same components of $\partial Q^{(0)}$ as those of the trajectory of x. Then by (3) the reflections of $T^t x$ and $T^t x'$ from each cylinder occur at points that are displaced relative to each other along a generator of the cylinder. This means that for every pair of colliding discs the vectors of their centre lines corresponding to $T^t x$ and $T^t x'$ coincide. In addition, the vectors normal to each cylinder at the points of reflection of $T^t x$ and $T^t x'$ coincide, so the velocity vectors of all discs at the instants of the first $\tilde{l}$ collisions also coincide. This leads to non-uniqueness of reconstruction of the coordinates of the discs at the instants of the first $\tilde{l}$ collisions, and this contradicts Proposition 1. The same reasoning goes over unchanged to the ball case.

§4. Local ergodicity

We first prove Theorems 1–3. The simplest is the proof of Theorem 3, since the corresponding entropy technique is sufficiently developed.

Let $B(x) \geqslant 0$ be an operator of the second fundamental form for an LUM of a semiscattering billiard system. We claim that the entropy of the automorphism T_1 is

$$h(T_1) = \int_{\mathfrak{M}_1} \log \det (I + \tau B(x))\, d\mu_1(x),$$

where $\tau > 0$ is the time interval until the next reflection. General formulae for the entropy of a flow or of an automorphism with invariant measurable foliations were obtained in [18]. A similar formula for flows with the property of full hyperbolicity that have singularities is discussed in [31].

We note that a family of operators $B(x)$ for an LSM and an LUM defines a decomposition of the tangent bundle of the phase space into invariant subbundles, and a formula of the above type is obtained with the help of the Pesin technique (see [13]). We do not dwell on this in more detail.

We now turn to the proof of Theorem 1. It is sufficient to prove the corresponding statement for T_1. From the conditions of the theorem $j(x) = 2n-3$ ($j(x) = 3n-4$ in the ball case) almost everywhere and therefore it is equal to half of the dimension of $\partial\mathfrak{M}^{(0)}$. Let $W_1^{(u)}(x)$ and $W_1^{(s)}(x)$ denote an LUM and an LSM of a point $x \in \partial\mathfrak{M}^{(0)}$.

For almost every x these $W_1^{(u)}(x)$, $W_1^{(s)}(x)$ belong to the same ergodic component as x. One can show that the families $\{W_1^{(u)}(x)\}$, $\{W_1^{(s)}(x)\}$ have the property of absolute continuity. In the present case it means the following. Each $W_1^{(u)}(x)$ is a smooth submanifold. We take an arbitrary subset $C \subset W_1^{(u)}(x)$ of positive measure and draw an LSM $W_1^{(s)}(y)$ through almost every point $y \in C$. The absolute continuity means that $\bigcup_{y \in C} W_1^{(s)}(y)$ is a set of positive measure in $\partial\mathfrak{M}^{(0)}$. The property of absolute continuity was first established for Anosov systems (see [2]). For a discontinuous system with singularities, to which the present case belongs, a detailed presentation of absolute continuity is in [31].

The reasoning of Hopf (see [18], [19]) shows that for almost every x the set $\bigcup_{y \in W_1^{(u)}(x)} W_1^{(s)}(y)$ belongs mod 0 to one ergodic component. Because of the absolute continuity this set has positive measure. We have thus proved that the ergodic components of T_1 and therefore of $\{T^t\}$ have positive measure.

We now investigate the K-property of T_1. The method employed below first appeared in [18]. We shall use the notion of π-partition, that is, the maximal partition with zero entropy (for T_1). We introduce global fibres

$$\overline{W}_1^{(u)}(x) = \bigcup T_1^n W_1^{(u)}(T_1^{-n}x), \quad \overline{W}^{(s)}(x) = \bigcup T_1^{-n} W_1^{(s)}(T_1^n x).$$

Then π does not exceed the measurable hull of the partition, whose elements include with almost every x the $\overline{W}_1^{(u)}(x)$, $\overline{W}_1^{(s)}(x)$ containing x. The latter partition is discrete because of the absolute continuity. Hence the π-partition is also discrete. But then the π-partition for $\{T^t\}$ is discrete too. Since this partition is invariant, π is a partition into ergodic components.

To prove Theorem 2 it is sufficient to establish that under the conditions of the theorem there is an open set where $j(x) = 2n-3$. Indeed, the set where $j(x) = 2n-3$ is open and invariant mod 0. Consequently, the previous reasoning is applicable to this set.

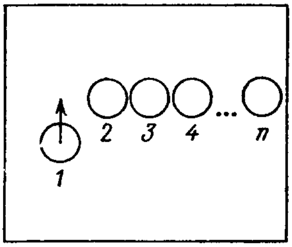


Fig. 1

We consider a system of n discs of radius $r \leqslant 1/2n$. They can be arranged as shown in Fig. 1: one moving disc with the velocity vector $v_1 = (0, 1)$ and $n-1$ stationary discs, whose centres lie on a horizontal straight line. We consider velocity vectors in the coordinate system connected with the second disc; the transition to the original coordinate system, in which $P = 0$, is trivial. Suppose that discs 1 and 2 collide in such a way that the vector of their centre line at the instant of collision is almost horizontal. Then the 2nd disc starts to move and there occurs a chain of collisions of the 2nd and 3rd discs, the 3rd and 4th discs, ..., the $(n-1)$st and n-th discs. It is easy to see that the velocity vector of the i-th disc $(i \geqslant 2)$, obtained by the disc immediately after its collision with the $(i-1)$st disc, has the form (α_i, β_i), where $|\beta_i| \ll \alpha_i$ and $\beta_i > 0$ for even i and $\beta_i < 0$ for odd i. The dynamics of this system in the past $(t < 0)$ is similar up to symmetry, that is, the vectors (α_i', β_i') are such that $|\beta_i'| \ll \alpha_i'$ and $\beta_i' > 0$ for odd i and $\beta_i' < 0$ for even i.

We prove Proposition 1 for this case, that is, we show that the coordinates of the discs can be uniquely reconstructed from the totality of all velocity vectors and directions of centre lines of the discs colliding in the $2(n-1)$ collisions (in the past and in the future) under consideration. We fix an initial position of the 2nd disc (this is equivalent to fixing the centre of gravity of the system). Then the direction of the centre line of the colliding pair (1, 2) uniquely determines the coordinates of the 1st disc. We reconstruct the coordinates of the 3rd disc, knowing the direction of the centre line of the colliding 2nd and 3rd discs. The set of possible positions of the centre of the 3rd disc forms a straight line parallel to the vector (α_2, β_2). Similarly, considering the collision of the 2nd and the 3rd discs in the past, we localize the position of the centre of the 3rd disc on a straight line parallel to the vector (α_2', β_2'). However, $\beta_2 > 0$, $\beta_2' < 0$, so these straight lines are not parallel, and this determines their unique point of intersection, which is the position of the centre of the 3rd disc. Continuing this reasoning, we reconstruct the coordinates of the 4th disc, ..., the n-th disc.

As we showed in §3, the relation $j(x) = 2n-3$ follows from Proposition 1, that is, this relation is satisfied in some neighbourhood of the point of the phase space that corresponds to the position of the discs in Fig. 1. This concludes the proof of Theorem 2.

Similar reasoning is applicable to the case of balls for the example shown in Fig. 1. The centres of all the balls are located in the same plane and to prove Proposition 1 an analysis similar to that presented above is carried out.

We now turn to a finer investigation of the structure of ergodic components. We have already mentioned the fundamental theorem of the theory of scattering billiards, which provides the possibility of establishing the uniqueness of the ergodic component. The first step in the proof of the fundamental theorem consists in showing that for points of general form there is a neighbourhood that belongs mod 0 to one ergodic component. Below we carry out a different proof of this statement, which is applicable to both semiscattering and scattering billiards. In §5 we use this proof to single out the case of three discs, for which one can prove the full ergodicity and the K-property.

For semiscattering billiards we consider a manifold $R \subset \partial\mathfrak{M}$ consisting of points of discontinuity of the map T_1, of singular points of the boundary $\partial\mathfrak{M}$, and of points of the set $\mathfrak{M}_2 = \{x \in \partial\mathfrak{M}\colon (n, v) = 0\}$. We assume that one of the following conditions is satisfied:

A. For almost every $x \in T_1^2 R$ the following relation is satisfied (in the sense of the intrinsic Riemannian metric on $T_1^2 R$):

$$\| ((I + \tau_n B_n) U_{n-1}^{-1} (I + \tau_{n-1} B_{n-1}) U_{n-2}^{-1} \ldots U_2^{-1} (I + \tau_1 B_1))^{-1} \| \xrightarrow[n \to \infty]{} 0,$$

where B_n is the curvature operator at the point $T^{t_n+0}x$ of the image of the flat fibre $\Sigma^{(0)}$ that contains the point $-x$ under the action of T^{t_n+0}. (A flat fibre is a fibre with a flat support $\widetilde{\Sigma}^{(0)}$.) The notations U_i, t_i, τ_i were introduced in §3. In other words, the flat fibre containing x expands unrestrictedly in all directions at πx under the action of T^t as $t \to \infty$ (see [34]).

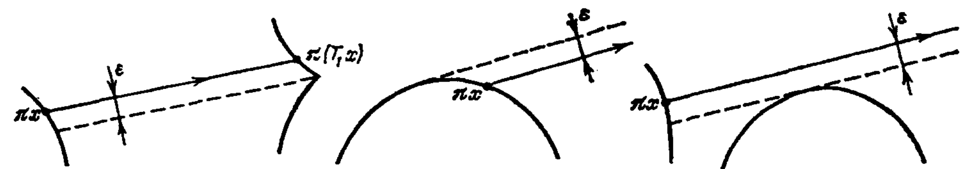
B. For almost every $x \in R$ (in the sense of the intrinsic Riemannian metric) there is an LSM $W_1^{(s)}(T_1^2 x)$ at $T_1^2 x$ relative to the derived automorphism T_1.

Condition A is weaker than B (we shall show this) and each of them is sufficient for the proof of Theorem 5 formulated below.

***Theorem* 5.** *Let $x \in \mathfrak{M}'$ be such that $B(x) > 0$, $B(-x) > 0$, where $-x = (q, -v)$ if $x = (q, v)$ and either* A *or* B *is satisfied. Then there is a neighbourhood $U(x) = \mathfrak{M}$ of x that belongs* mod 0 *to one ergodic component.*

Proof. We first make the following remark. In the case of semiscattering billiards in bounded domains of Euclidean space the length of a free path is uniformly bounded, and the number of manifolds of discontinuity for T_1 is finite. This can also be achieved in the case of domains belonging to a torus. Namely, let a torus be generated by pasting together suitable faces of a parallelepiped K. We consider ∂Q after the addition of ∂K and continue to denote by T_1 the derived automorphism corresponding to this expansion of the boundary (see §2). An invariant measure μ_1 of T_1 at points $q \in \partial K$ has the density $d\mu_1(q, v) = |(n(q), v)| dq\, d\omega_S$, where $n(q)$ is the normal vector to ∂K.

Let $z(x)$ denote the maximal diameter of the base of a cylindrical neighbourhood in Q of the segment of the trajectory that connects x and $T_1^{-1} x$ and terminates at two regular components of the (expanded) boundary ∂Q. For points $x \in \mathfrak{M}$ we put $z(x) = z(\pi_1 x)$.



a b c

Fig. 2

Lemma 2. $\mu_1\{x \in \mathfrak{M}_1 : z(x) < \varepsilon\} \leqslant \mathbf{const}\,(Q)\varepsilon$ *for all* $\varepsilon > 0$.

Proof. It is sufficient to consider only sufficiently small ε. For such ε there are three possibilities.

1°. One of the points $\pi(x)$, $\pi(T_1^{-1}x)$ lies in the ε_1-neighbourhood of the set of singular points Q_{00} of the boundary $\varepsilon_1 = \varepsilon/\cos\varphi$, $\cos\varphi = (n(q), v)$ (Fig. 2a).

Q_{00} consists of finitely many smooth submanifolds of codimension 1 in ∂Q. Hence the measure of its ε_1-neighbourhood in ∂Q does not exceed $\mathbf{const}(Q)\varepsilon_1$ (see [27]). Then the μ_1-measure of such points does not exceed

$$\mathbf{const}\,(Q)\int_S \frac{\varepsilon}{\cos\varphi}\cos\varphi\, d\omega_S = \mathbf{const}\,(Q)\,\varepsilon.$$

2°. $\varphi > \frac{\pi}{2} - \mathbf{const}(Q)\sqrt{\varepsilon}$ (Fig. 2b). Then $\cos\varphi < \mathbf{const}(Q)\sqrt{\varepsilon}$, and the μ_1-measure of such points does not exceed

$$\int_{\partial Q}\ \int_{\cos\varphi < \mathrm{const}\,(Q)\sqrt{\varepsilon}} \mu_1(d\omega, dq) \leqslant \mathbf{const}\,(Q) \int_{\frac{\pi}{2} - \mathrm{const}\,(Q)\sqrt{\varepsilon}}^{\pi/2} \cos\varphi\, d\varphi \leqslant \mathbf{const}\,(Q)\cdot\varepsilon.$$

3°. Part of the trajectory in question lies in the ε-neighbourhood of a regular boundary component and does not intersect it (Fig. 2c). Then the μ_1-measure of such points does not exceed $\mathbf{const}(Q)S(Q)\varepsilon$, where $S(Q)$ is the area of ∂Q.

The lemma is proved.

We return to the initial point x with $B(x) > 0$ and $B(-x) > 0$. It is shown in [23] that $j(y) = j(-y) \equiv d-1$ in a sufficiently small neighbourhood $U(x)$ and that for almost every $y \in U(x)$ there are $(d-1)$-dimensional LSM $W^{(s)}(y)$ and LUM $W^{(u)}(y)$. It is sufficient to prove the theorem for a sufficiently small neighbourhood $U_1(x_1) \subset \mathfrak{M}_1$ of the point $x_1 = \pi_1(x)$. The projections $W^{(s)}$, $W^{(u)}$ onto $\mathfrak{M}_1$ provide the LSM and the LUM $W_1^{(s)}(y_1)$, $W_1^{(u)}(y_1)$ for $y_1 \in U_1(x_1)$ relative to T_1. Their dimensions are equal to $d-1$, while $\dim U_1(x_1) = 2d-2$. From the continuity of $B(x)$ on $\mathfrak{M}'$ and Theorem 4 it follows that the spaces $\mathcal{T}_y W_1^{(s)}$, $\mathcal{T}_y W_1^{(u)}$ tangent to $W_1^{(s)}$, $W_1^{(u)}$ depend continuously on $y \in U_1(x_1) \cap \mathfrak{M}'$. Taking $U_1(x_1)$ sufficiently small we can assume that $\mathcal{T}_y W_1^{(s)}$ $(\mathcal{T}_y W_1^{(u)})$ are sufficiently close to each other for different y. The inequalities $B(x) > 0$, $B(-x) > 0$ also mean that $\mathcal{T}_y W_1^{(s)}$, $\mathcal{T}_y W_1^{(u)}$ are mutually transversal and their direct sum is $\mathcal{T}_y \mathfrak{M}_1$.

We introduce a small parameter δ that will later tend to zero. We consider a family of open covers $U_1(x_1)$ that depend on δ with the following properties. An element G of each cover is a "parallelogram", that is, the image of a $(2d-2)$-dimensional cube under the linear map $\mathbf{R}^{2d-2} \to \mathfrak{M}_1$, so that for a certain point $y \in G$ the tangent spaces $\mathcal{T}_y W_1^{(s)}$, $\mathcal{T}_y W_1^{(u)}$ are parallel to the corresponding faces of G, while the length of each edge of G is equal to δ. The covers $\{G_i^{(\delta)},\ 1 \leqslant i \leqslant I(\delta)\}$ that we shall consider are such that:

a) the distances between the tangent spaces to points on the faces of the same name of different parallelograms $G_i^{(\delta)}$ are sufficiently small, while their centres form a finite set that is sufficiently close to a certain lattice depending on the size of $U_1(x_1)$;

b) if $G_i^{(\delta)} \cap G_j^{(\delta)} \neq \varnothing$ for some $i \neq j$, then $\mu(G_i^{(\delta)} \cap G_j^{(\delta)}) \geqslant c_0\delta^{2d-2}$, where c_0 does not depend on δ;

c) each point $y \in U_1(x_1)$ belongs to at most 2^{2d-1} different parallelograms.

Let $G_i^{(\delta)}$ be a parallelogram from the cover, and $y \in G_i^{(\delta)}$ a marked point. There are 2^{2d-1} $(d-1)$-dimensional faces of $G_i^{(\delta)}$ that are parallel to $\mathcal{T}_y W_1^{(u)}$. We call these u-leading faces. Correspondingly, there are 2^{2d-1} $(d-1)$-dimensional faces parallel to $\mathcal{T}_y W_1^{(s)}$, which we call s-leading faces.

***Lemma* 3.** *For every $\delta \leqslant \delta_0(x)$ one can split the set of parallelograms from $\{G_i^{(\delta)}\}$ into two groups $\mathfrak{A}_0^{(\delta)}$ and $\mathfrak{A}_1^{(\delta)}$ so that:*

a) *for $G_i^{(\delta)} \in \mathfrak{A}_1^{(\delta)}$ we take any of its u-leading faces and its $c_1\delta$-neighbourhood $U^{(u)}$, where c_1 does not depend on δ, but can be made arbitrarily small by decreasing $U_1(x_1)$; we consider $z \in U^{(u)}$ such that $\partial(W_1^{(u)}(z) \cap G_i^{(\delta)})$ belongs to the union of s-leading faces. Then the measure of such z's is positive. A similar statement is true for every s-leading face;*

b) *the measure of the union of all parallelograms from $\mathfrak{A}_0^{(\delta)}$ is equal to $\delta\varphi_1(\delta)$, where $\varphi_1(\delta) \to 0$ as $\delta \to 0$.*

The detailed derivation of Theorem 5 from Lemma 3 is given in [20]. Here we give only a sketch of the proof. The proof is based on an idea of Hopf [23], consisting in the fact that all LSM's and LUM's belong mod 0 to one ergodic component. Indeed, for every continuous function f on $\mathfrak{M}_1^+$ the time averages

$$f^{(+)}(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T_1^i y)$$

are constant on $W_1^{(s)}$, since $\operatorname{diam}(T_1^i W_1^{(s)}) \to 0$ as $i \to \infty$, and $\mathfrak{M}_1^+$ is a compactum. Similarly

$$f^{(-)}(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T_1^{-i} y)$$

is constant on each $W_1^{(u)}$. Therefore both $W_1^{(s)}(x)$, $W_1^{(u)}(x)$, and their union, also belong mod 0 to one ergodic component.

As shown in [20], it follows from Lemma 3 that one can find a set of parallelograms $\mathfrak{A}_2^{(\delta)} \subset \mathfrak{A}_1^{(\delta)}$ whose union is connected, while its measure tends to $\mu_1(U_1(x_1))$ as $\delta \to 0$. Hence, for any pair of points x', x'' that belong to a subset of full measure in $U_1(x_1)$ there are LSM's $W_1^{(s)}(x')$ and $W_1^{(s)}(x'')$ of sizes δ' and δ'' respectively, and for some $\delta < \min\{\delta', \delta''\}$ both x' and x'' are included in parallelograms from $\mathfrak{A}_2^{(\delta)}$. By Property a) of Lemma 2 we can find a chain of LSM's and LUM's $W_{1,1}^{(s)}, W_{1,2}^{(u)}, \ldots, W_{1,k}^{(u)}$ so that $x' \in W_{1,1}^{(s)}$, $x'' \in W_{1,k}^{(u)}$ and $W_{1,i}^{(s)} \cap W_{1,i+1}^{(u)} \neq \varnothing$ for all $i = 1, 2, \ldots, k-1$. The theorem is proved.

Proof of Lemma 3. The first step, as in [34], consists in choosing "too bad" parallelograms and including them in $\mathfrak{A}_0^{(\delta)}$. The map T_1 is piecewise continuous. For each $k \in \mathbf{Z}$ the power T_1^k is discontinuous on the union of finitely many submanifolds with boundaries of codimension 1 (we recall that we expanded the boundary ∂Q by adding ∂K to it; see above), which are images of the submanifolds $Q_{00} \times S$ and $\mathfrak{M}_2$ under the action of T_1^j ($j = 0, -1, \dots, -k+1$). Let $\mathfrak{A}_{00}^{(\delta)}$ be the set of those parallelograms $G_i^{(\delta)}$ that intersect at least two manifolds of discontinuity of T_1^k for $k \leqslant F(\delta)$. One can choose $F(\delta)$ in such a way that $F(\delta) \to \infty$ as $\delta \to 0$, and the measure of $\mathfrak{A}_{00}^{(\delta)}$ is equal to $\delta\varphi_2(\delta)$, where $\varphi_2(\delta) \to 0$ as $\delta \to 0$ (this is explained in more detail in [33]). The set $\mathfrak{A}_{00}^{(\delta)}$ is included in $\mathfrak{A}_0^{(\delta)}$.

Further, dim $J_0(y) \equiv 0$ in $U(x)$. Hence it follows from (4) that every point $y \in U(x) \cap \mathfrak{M}'$ is a point of local maximum of the function $l_0(y)$ introduced in §3. We can assume that $l_0(y) \leqslant l_0(x)$ for all $y \in U(x) \cap \mathfrak{M}'$. We denote by $l_1(x)$ the number of reflections of the trajectory of x from the expanded boundary ∂Q up to the $l_0(x)$th reflection from the unexpanded boundary (inclusive). Let $0 < t_1 < t_2 < \dots$ be the instants of the reflections of the trajectory of x from the expanded boundary and suppose that $t^* \in (t_{l_1(x)}, t_{l_1(x)+1})$. Let $U(x)$ be so small that T^{t^*} is smooth on $U(x)$. Then the support $\tilde{\Sigma}$ of any convex LM Σ that contains $-T^{t^*}y$ for $y \in U(x)$ expands in all directions under the action of $\pi \circ T^{-t^*} \circ \pi^{-1}$, where T^{-t^*} acts on Σ. If $D = D(y, t^*, \Sigma)$ is the differential of the map $\pi \circ T^{-t^*} \circ \pi^{-1}|_{\tilde{\Sigma}}$ at the point $\pi(-T^{t^*}y)$, then

$$\|D^{-1}\| \leqslant \lambda < 1, \tag{6}$$

where λ depends only on x, but not on y, t^*, Σ (see the proof in [24]). In other words, the support of any convex LM that is included in $-T^{t^*}U(x) = \{-y\colon y \in T^{t^*}U(x)\}$ expands under the action of $\pi \circ T^{-t^*} \circ \pi^{-1}$ in every direction by a factor at least $\Lambda = \lambda^{-1} > 1$. We choose $U_1(x_1)$ so small that $T^t y \notin U(x)$ for all $y \in U(x)$, $0 < t \leqslant t_{l_1(x)}$. (In the case of a periodic point x such a choice is also possible, since $t_{l_1(x)}$ does not exceed the period of the trajectory of x.)

Let $y \in \mathfrak{M}'$ and $t > 0$, $\delta > 0$. We consider all possible convex fibres Σ containing the point $-T^t y$ on which T^t is smooth, and are such that the distance from $\pi(-T^t y)$ to the most remote point of the boundary $\partial\tilde{\Sigma}$ of the support in the intrinsic metric of $\tilde{\Sigma}$ does not exceed δ. We denote by $\varkappa_{t,\delta}(y)$ the minimal expansion factor in any direction of the supports of such fibres at any of their points under the action of T^t. Note that $\varkappa_{t,\delta}(y)$ increases monotonically as $\delta \to 0$ and has the limit $\varkappa_{t,0}(y)$, which is equal to the minimal eigenvalue of the operator (see [34])

$$(I + \tau_1 B_1^{(l)}) U_1 (I + \tau_2 B_2^{(l)}) U_2 \dots U_{l-1} (I + \tau_l B_l^{(l)}), \tag{7}$$

where $B_i^{(l)}$ is the curvature operator at $-T^{t_i - 0}y$ of the image of the flat fibre $\Sigma(i)$ that contains $-T^t y$ (we recall that a flat fibre means a fibre with flat support $\tilde{\Sigma}^{(t)}$) under the action of T^{t-t_i+0}; the notations U_i, t_i, τ_i were introduced in §3, and l is the number of reflections of $T^s y$ from ∂Q for $0 < s < t$.

We consider the set

$$U_1^{(0)} = \{y \in U_1(x_1):\ z(T^t y) \geqslant c_2 \delta \varkappa_{t,\, c_2\delta}(y) \ \text{ for all } \ t > 0\},$$

where the value of the constant c_2 is defined below. We claim that for every $y \in U_1^{(0)}$ the size of $W_1^{(s)}(y)$ in every direction is at least $\sqrt{d\delta}$. Let $r(y)$, $y \in \mathfrak{M}$, be the distance from $\pi(y)$ to $\partial\tilde{W}^{(s)}(y)$, where $\tilde{W}^{(s)}(y)$ is the support of $W^{(s)}(y)$, and the distance is taken in the sense of the induced Riemannian metric on $W^{(s)}(y)$. We put $r(y) = 0$ if there is no LSM for y.

***Lemma* 4.** *If $y \in U_1^{(0)}$, then $r(y) \geqslant c_2\delta$.*

Proof. Let $\tilde{\Sigma}^{(t)}$ be the $(d-1)$-dimensional disc of radius $\min\{c_2\delta,\ z(T^t y)\}$ in Q with centre at $\pi(T^t y)$ that is orthogonal to the velocity vector v_t of $T^t y$. We construct an LM $\Sigma^{(t)}$ with support $\tilde{\Sigma}^{(t)}$ and velocity vectors equal to $-v_t$. The map $T^t|_{\Sigma^{(t)}}$ is piecewise smooth. We denote by $\Sigma_0^{(t)}$ the domain in $\Sigma^{(t)}$ containing $-T^t y$ on which T^t is smooth. The LM $\Sigma_0^{(t)}$ and its images $\Sigma_{0,s}^{(t)} = T^s\Sigma_0^{(t)}$ are convex for $0 \leqslant s \leqslant t$, therefore their supports do not contract in any direction as s increases. Moreover, by the definition of $\varkappa_{t,\delta}$ the support $\tilde{\Sigma}_0^{(t)}$ expands at least $\varkappa_{t,\, c_2\delta}(y)$ times in every direction under the action of T^t.

Let $r_t(y)$ be the distance from $\pi(y)$ to the nearest point $q^{(t)}$ of the boundary $\partial\tilde{\Sigma}_{0,t}^{(t)}$ (in the sense of the Riemannian metric on $\tilde{\Sigma}_{0,t}^{t}$). Suppose that $r_t(y) < c_2\delta$ for some $t > 0$. If $q^{(t)}$ is obtained from some point of $\partial\tilde{\Sigma}^{(t)}$, then $z(T^t y) \leqslant c_2\delta$ and $r_t(y) \geqslant z(T^t y)\varkappa_{t,\, c_2\delta}(y) \geqslant c_2\delta$. Otherwise $q^{(t)}$ is the support of some point of $\partial\Sigma_{0,t}^{t}$, which goes into a singular point of the boundary $y' \in \partial\mathfrak{M}$ or into a point $y' \in \mathfrak{M}_2$ under the action of T^s for some $s \in (0, t)$. By the definition of $z(x)$ we have $z(T^s y) \leqslant \operatorname{dist}(\pi y', \pi T^s y) \leqslant c_2\delta$, and therefore $r_t(y) \geqslant z(T^s y)\varkappa_{s,\, c_2\delta}(y) \geqslant c_2\delta$. Thus, in all cases $r_t(y) \geqslant c_2\delta$, which proves the lemma, since $W^{(s)}(y)$ is the limit of the fibres $-\Sigma_{0,t}^{t}$ as $t \to \infty$.

The map π_1 is smooth at x. Hence c_2 can be chosen in a way that the size of the support of $W_1^{(s)}(y) = \pi_1 W^{(s)}(y)$ is at least $\sqrt{d\delta}$ in every direction. Consequently, all $y \in U_1^0$ also have LSM's of the necessary size.

We now investigate the complement $U_0 = U_1(x_1) \setminus U_1^{(0)}$. We put $\varkappa_{n,\delta}(y) = \varkappa_{t_{n-1},\,\delta}(y)$, where t_n is the instant of the n-th reflection of $T^t y$ from the expanded boundary ∂Q. It is clear that $U_0 = U_{0,1} \cup U_{0,2} \cup \cdots$, where

$$U_{0,n} = \{y \in U_1(x_1):\ z(T_1^n y) \leqslant c_2\delta\varkappa_{n,\, c_2\delta}^{-1}(y)\},$$

and $U_{0,n} \subset U_{0,n,0} \cup U_{0,n,1} \cup \cdots$, where

$$U_{0,n,m} = \{y \in U_{0,n}:\ \log_\Lambda \varkappa_{n,\, c_2\delta}(y) \in [m,\ m+1)\},$$

where the constant $\Lambda > 1$ was introduced in the note to formula (6). Therefore

$$(8) \qquad \mu_1(U_0) \leqslant \sum_{n=2}^{\infty} \sum_{m=0}^{\infty} \mu_1(U_{0,n,m}) = \sum_{m=0}^{\infty} \sum_{n=2}^{\infty} \mu_1(T_1^{n-1} U_{0,n,m}).$$

***Lemma* 5.** *For every $m \geqslant 0$, $n_1 \geqslant 2$, $n_2 \geqslant 2$, $n_1 \neq n_2$ the sets $T_1^{n_1-1} U_{0,n_1,m}$ and $T_1^{n_2-1} U_{0,n_2,m}$ do not intersect.*

Proof. Let $n_1 < n_2$ and $y \in T_1^{n_1-1} U_{0,n_1,m} \cap T_1^{n_2-1} U_{0,n_2,m}$. Then $y' = T_1^{-n_1+1} y \in U_{0,n_1,m} \subset U_1(x_1)$ and $y'' = T_1^{-n_2+1} y \in U_{0,n_2,m} \subset U_1(x_1)$. Note that $y' = T^{t'} y$ and $y'' = T^{t''} y$ for some t', $t'' > 0$. Each convex fibre Σ of size not exceeding $c_2\delta$ that contains $-y$ expands at least $\varkappa_{n_1, c_2\delta}(y')$ times under the action of $T^{t'}$. However, by (6), the fibre $T^{t'}\Sigma$ expands at least Λ times under the action of $T^{t''-t'}$, therefore

$$\varkappa_{n_2, c_2\delta}(y'') \geqslant \varkappa_{n_1, c_2\delta}(y') \cdot \Lambda,$$

which contradicts the definition of the sets $U_{0,n,m}$. Lemma 5 is proved.

Lemma 5 enables us to transform (8):

$$(9) \qquad \sum_{n=2}^{\infty} \mu_1(T_1^n U_{0,n,m}) = \mu_1\left(\bigcup_{n=2}^{\infty} T_1^n U_{0,n,m}\right) \leqslant$$
$$\leqslant \mu_1\{y \in \partial\mathfrak{M}:\ z(y) \leqslant c_2\delta\lambda^m\} \leqslant c_3(x_1)\,\delta\lambda^m,$$

the latter inequality follows from Lemma 2. From (8) and (9) we obtain

$$(10) \qquad \mu_1(U_0) \leqslant \sum_{m=0}^{\infty} c_3(x_1)\,\delta\lambda^m = c_4(x_1)\,\delta.$$

We put $U_{0\alpha} = \bigcup_{m=0}^{F(\delta)} U_{0,m}$ and $U_{0\omega} = \bigcup_{n=F(\delta)+1}^{\infty} U_{0,n}$, where $F(\delta)$ was introduced at the beginning of the proof of Lemma 2. We now prove that

$$(11) \qquad \mu_1(U_{0\omega}) \leqslant \delta\varphi_3(\delta),$$

where $\varphi_3 \to 0$ as $\delta \to 0$. One of the additional conditions, either A or B, is needed for the proof.

***Lemma* 6.** *If for every $m \geqslant 0$ and for any function $F_1(\delta) \to \infty$ as $\delta \to 0$*

$$\sum_{n=F_1(\delta)}^{\infty} \mu_1(U_{0,n,m}) = o(\delta) \quad as \ \delta \to 0,$$

then (11) *is true.* (*Note that the sets $U_{0,n,m}$ depend on the parameter δ.*)

Proof. Let $\varepsilon > 0$. By (9) there is an m_0 such that

$$\sum_{m=m_0}^{\infty} \sum_{n=2}^{\infty} \mu_1(U_{0,n,m}) \leqslant \frac{\varepsilon\delta}{2}.$$

According to the conditions of Lemma 6 there are numbers $N(m)$ and $\delta_0(m)$ for every $m = 0, 1, \ldots, m_0 - 1$ such that for all $\delta < \delta_0(m)$

$$\sum_{n=N(m)}^{\infty} \mu_1(U_{0,n,m}) \leqslant \frac{\varepsilon\delta}{2m_0}.$$

Then for all δ such that $\delta < \delta_0(m)$ and $F(\delta) > N(m)$ for every $m = 0, 1, \ldots, \ldots, m_0 - 1$ we have

$$\sum_{n=F(\delta)+1}^{\infty} \mu_1(U_{0,n}) \leqslant \sum_{m=0}^{\infty} \sum_{n=F(\delta)+1}^{\infty} \mu_1(U_{0,n,m}) \leqslant \frac{\varepsilon\delta}{2} + m_0 \frac{\varepsilon\delta}{2m_0} = \varepsilon\delta,$$

and this proves Lemma 6.

Suppose that the conditions of Lemma 6 are not satisfied, that is, there are numbers $m_1 \geqslant 0$, $\varepsilon_1 > 0$ and a function $F_1(\delta) \to \infty$ as $\delta \to 0$ such that

$$\sum_{n=F_1(\delta)}^{\infty} \mu_1(U_{0,n,m_1}) \geqslant \varepsilon_1\delta. \tag{12}$$

We consider the manifold R and the set W of "degenerate trajectories"

$$W = \{x \in \partial\mathfrak{M}:\ \varkappa_{0,t}(-T^t x) = 1 \quad \text{for all} \quad t > 0\},$$

that is, the set of points x such that the support of the flat fibre containing x does not expand (and does not bend) under the action of T^t for all $t > 0$ in one of the directions at $\pi(x)$. Clearly W is semi-invariant: $T_1^n W \subset W$ for $n > 0$. Since $B(x) > 0$ almost everywhere on $\mathfrak{M}$, it follows that $\mu_1(W) = 0$ and W is nowhere dense in $\partial\mathfrak{M}$. From Condition A it follows immediately that the set $T_1^2 R \cap W$ has zero Riemannian volume on R. This follows also from Condition B, since otherwise the points $x \in T_1^2 R \cap W$ form, together with their LSM's, a set of positive μ_1-measure in $\partial\mathfrak{M}$ (because of the absolute continuity of transversal foliations; see [2], [31]), but this set is attracted to W as $n \to \infty$, which contradicts the condition $\mu_1(W) = 0$.

Let us derive A from B. Let $x \in T_1^2 R \setminus W$ and let $V(x)$ be a small neighbourhood of it in $\mathfrak{M}$. The union of the LSM's of the points $y \in (T_1^2 R \setminus W) \cap V(x)$ forms a set of positive μ_1-measure in $\partial\mathfrak{M}$ because of the absolute continuity of transversal foliations. By the Poincaré recurrence theorem [2] almost every point of this union (in the sense of μ_1) returns to the neighbourhood $V(x)$ infinitely many times, so $\varkappa_{t,0}(-T^t y) \to \infty$ as $t \to \infty$. Consequently, $\varkappa_{t,0}(-T^t y) \to \infty$ as $t \to \infty$ for almost every $y \in T_1^2 R$ (in the sense of the Riemannian metric on $T_1^2 R$), and this is equivalent to A according to (7).

Hence, from A or B it follows that $\varkappa_{t,0}(-T^t y) \to \infty$ as $t \to \infty$ for almost every $y \in T_1^2 R$ in the sense of the Riemannian metric on $T_1^2 R$. Thus, for such y there is a neighbourhood $V_1(y)$ in $\mathfrak{M}$ such that $\varkappa_{t,0}(-T^t y') > \Lambda^{m_1+2}$ for $t > t_1(y)$ for all $y' \in V_1(y)$. Further, there is a smaller neighbourhood $V_2(y) \subset V_1(y)$ such that $\varkappa_{t,\delta}(-T^t y') > \Lambda^{m_1+2}$ for $t > t_2(y)$ for all $y' \in V_2(y)$ and $\delta < \delta_2(y)$. This means that the set $T_1^{n-1} U_{0,n,m_1}$ does not intersect

$T_1^s V_2(y)$ for any $\delta < \delta_2(y)$ and $n \geqslant n_2(y)$ (n_2 depends on t_2). Hence it follows that for every $\varepsilon_0 > 0$ there are numbers $\delta_0 > 0$ and $n_0 > 0$ and a neighbourhood $V_0(R)$ of R in $\partial\mathfrak{M}$ such that

$$\mu_1 (V_0 (R) \cap (\bigcup_{n=n_0}^{\infty} T_1^{n-1} U_{0,\,n,\,m_1})) < \varepsilon_0 \delta$$

for all $\delta < \delta_0$, which contradicts (12). Relation (11) is thus proved. Now we consider in more detail the points $y \in \{r(y) < c_2\delta\} \setminus U_{0\omega}$.

As before, $0 < t_1 < t_2$ ... means a sequence of instants of reflections from the boundary. For $t > t_{F(\delta)}$ and $s > 0$ we take the LM $\Sigma_{0,\,s}^{(t)}$ introduced in the proof of Lemma 4. For $y \notin U_{0\omega}$ the size of $\pi(\Sigma_{0,\,t-t_{F(\delta)}}^{(t)})$ in every direction is at least $c_2\delta\varkappa_{F(\delta),\,c_2\delta}^{-1}(y)$. The proof of this statement is completely analogous to the proof of Lemma 3. Therefore $\Sigma_{0,\,t}^{(t)}$ interesects the manifold of discontinuity of $T_{0,\,t}^{tF(\delta)}$ for all t, while $\pi_1\Sigma_{0,\,t}^{(t)}$ intersects the manifold of discontinuity of $T_1^{F(\delta)}$, and the distance between the latter and y in the metric of $\pi_1\Sigma_{0,\,t}^{(t)}$ is less than $\sqrt{d\delta}$. Because of the continuity of T_1 and its powers at x_1 we can choose $U_1(x_1)$ so small that the distance[1] between the spaces tangent to $\pi_1\Sigma_{0,\,t}^{(t)}$ and to the corresponding manifold of discontinuity is less than any preassigned $\varepsilon_1 > 0$. Then the distance between y and the corresponding manifold of discontinuity for $T_1^{F(\delta)}$ is at most $\varepsilon_2\delta$, where $\varepsilon_2 = \varepsilon_2(\varepsilon_1)$ can be made arbitrarily small if $U_1(x_1)$ is sufficiently small. Thus all points $y \in \{r(y) < c_2\delta\} \setminus U_{0\omega}$ lie in the $\varepsilon_2\delta$-neighbourhood of manifolds of discontinuity of $T_1^{F(\delta)}$.

Now we are able to construct the necessary sets $\mathfrak{A}_1^{(\delta)}$ and $\mathfrak{A}_0^{(\delta)}$ explicitly. Let the parallelogram $G_i^{(\delta)}$ intersect at most one manifold of discontinuity of $T_1^{\pm F(\delta)}$, which we denote by $\Sigma^{(i)}$. We put $G_i^{(\delta)} \subset \mathfrak{A}_1^{(\delta)}$ if

$$\mu_1 (G_i^{(\delta)} \cap U_1^{(o'')}) \geqslant (1 - \varepsilon_3)\mu_1(G_i^{(\delta)}).$$

Here ε_3 is the constant determined by c_1 in Condition a) of Lemma 2 for which the condition holds. We choose $U_1(x_1)$ and consequently ε_2 so small that the total measure of $\delta\varepsilon_2$-neighbourhoods of $\partial G_i^{(\delta)}$ and of the intersection of $\Sigma^{(i)}$ with $G_i^{(\delta)}$ does not exceed $\frac{1}{2}\varepsilon_3\mu_1(G_i^{(\delta)})$ for arbitrary δ and $G_i^{(\delta)}$.

If $G_i^{(\delta)} \notin \mathfrak{A}_1^{(\delta)}$, then either $G_i^{(\delta)}$ intersects at least the two manifolds of discontinuity of $T_1^{\pm F(\delta)}$, that is, $G_i^{(\delta)} \in \mathfrak{A}_{00}^{(\delta)}$, or $\mu_1(U_{0\omega} \cap G_i^{(\delta)}) \geqslant \frac{1}{2}\varepsilon_3\mu_1(G_i^{(\delta)})$. According to Property c) of the covers $\{G_i^{(\delta)}\}$ and the estimate (9) the measure of such parallelograms is equal to $\delta\varphi_4(\delta)$, where $\varphi_4(\delta) \to 0$ as $\delta \to 0$. Lemma 3 is proved, and so is Theorem 5.

---

[1] Here the distance between the tangent spaces means the following: we consider the intersection of each space with the unit sphere; further, for the points of the first of the resulting intersections we take the distance to the closest point of the second intersection and then consider the maximum of the two distances.

Theorem 5 permits the following generalization. Suppose that $x \in \mathfrak{M}_1$ belongs to exactly one manifold of discontinuity of T_1^k, $k \neq 0$. Without loss of generality we can assume that x belongs to the manifold of discontinuity Σ_{k_0} of the transformation $T_1^{k_0}$, $k_0 > 0$, and x is a non-singular point of Σ_{k_0}. Then Σ_{k_0} divides the neighbourhood $U(x) \subset \mathfrak{M}_1$ into two parts U_1 and U_2, and in this case we can define operators $B^{(1)}(x)$, $B^{(2)}(x)$ as the limits of $B(y)$ as $y \to x$, remaining either in U_1 or in U_2. Suppose that $B^{(i)}(-x) > 0$, $B^{(i)}(x) > 0$ for $i = 1, 2$.

***Theorem* 5′.** *Under the above conditions a sufficiently small neighbourhood of x belongs* mod 0 *to one ergodic component.*

Proof. Let $U(x)$ be so small that $B(\pm y) > 0$ for all $y \in U(x) \cap \mathfrak{M}'$. It easily follows from the proof of Theorem 5 that U_1 and U_2 belong mod 0 to one ergodic component.

In the proof of Theorem 5 we analysed the properties of the LUM $W^{(u)}(y)$. We showed that the set of points $y \in U(x)$ that do not have LUM $W^{(u)}(y)$ of size δ splits into two parts. The measure of the first part is $\delta\varphi(\delta)$, where $\varphi(\delta) \to 0$ as $\delta \to 0$. The second part lies in the $\varepsilon\delta$-neighbourhood of the manifolds of discontinuity of $T_1^{-F(\delta)}$, where $F(\delta) \to \infty$ as $\delta \to 0$, while ε can be made arbitrary small if $U(x)$ is sufficiently small.

In a small neighbourhood $U(x)$ the manifold Σ_{k_0} transversally intersects the manifolds of discontinuity of T_1^n when $n < 0$, since Σ_{k_0} is fibred into strictly convex LM's, while the other manifolds of discontinuity are fibred into strictly concave LM's. Hence at the point of intersection the corresponding tangent spaces generate the entire tangent space corresponding to this point. Therefore $F(\delta)$ can tend to infinity so slowly that the measure of the intersection of the δ-neighbourhood of Σ_{k_0} and the δ-neighbourhoods of the manifolds of discontinuity of $T_1^{-F(\delta)}$ has the form $\delta\varphi_4(\delta)$, where $\varphi_4(\delta) \to 0$ as $\delta \to 0$. But then, for sufficiently small δ, Σ_{k_0} intersects a set of positive measure consisting of LUM's that connect U_1 and U_2. Theorem 5′ is proved.

Note that Theorems 5 and 5′ enable us to prove the ergodicity of scattering billiards in bounded d-dimensional domains of Euclidean space and on a d-dimensional torus, $d \geqslant 2$. Indeed, for such billiards every point of $\partial\mathfrak{M}$ (except a finite set of singular points) is reflected from the scattering components of ∂Q infinitely many times. Hence A is always satisfied for such billiards and so Theorems 5 and 5′ are true. Consequently, every point of $\partial\mathfrak{M}$, except those whose trajectories $T_1^n x$ intersect the manifold of discontinuity of T_1 twice for $n = 0, \pm 1, \pm 2, \ldots$, has a neighbourhood that lies almost entirely in one ergodic component of T_1. Hence these points form a set M^+ whose complement consists of countably many submanifolds of codimension 2 in $\partial\mathfrak{M}$, that is, M^+ is linearly connected. Any two points of M^+ can be joined by a path that lies in M^+, and because of compactness the whole path can be covered by finitely many open sets, each of which lies almost entirely in one of the ergodic components of T_1, so their union also lies almost entirely in one ergodic component of T_1. This proves the ergodicity of T_1.

§5. Ergodicity of certain systems of three discs

We note firstly that all the statements of §§2-4 are true for systems of discs of various radii—appropriate changes should be made only in the formula (1) for cylinders.

We consider the system of three discs of radii $r_1 = r_2 = 1/8$ and $r_3 = 1/2$, whose centres are located at the points $O_1 = (3/8, 1/2)$, $O_2 = (5/8, 1/2)$, $O_3 = (0, 0)$ (Fig. 3). In this position the discs simultaneously touch each other at five points A_1–A_5. We shall study the motion of the discs for values of the parameters r_i and O_i sufficiently close to those mentioned above.[(1)] For such values of these parameters the centres of the moving discs are always located in a small neighbourhood of the initial positions, and the motion is actually reduced to the transformation of velocity vectors. We claim that for a sufficiently small neighbourhood of the values r_i, O_i in the parameter space the system of three discs does not have degeneracies in the phase space, that is, $W = \emptyset$ in the notation of §4, Theorem 5. For this it is sufficient to show that $j(x) \equiv 3$ everywhere in the phase space.

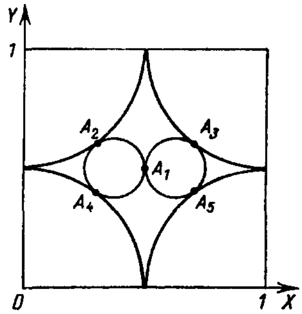


Fig. 3.

We introduce a coordinate system $(x_1, y_1, x_2, y_2, x_3, y_3)$ in the configuration space Q, where x_i, y_i are the coordinates of the i-th disc. We consider an arbitrary point x of the phase space and its trajectory $\{T^t x\}$. We denote the velocity vector of the i-th disc after the m-th collision by $(u_i^{(m)}, v_i^{(m)})$. If p_m and q_m are the numbers of the discs that participate in the m-th collision, then the coordinates of the vector e_m introduced in §3 are given by

$$p_m = 1, \quad q_m = 2: \ (v_1^{(m)} - v_2^{(m)},\ u_2^{(m)} - u_1^{(m)},\ v_2^{(m)} - v_1^{(m)},\ u_1^{(m)} - u_2^{(m)},\ 0,\ 0),$$
$$p_m = 1, \quad q_m = 3: \ (v_1^{(m)} - v_3^{(m)},\ u_3^{(m)} - u_1^{(m)},\ 0,\ 0,\ v_3^{(m)} - v_1^{(m)},\ u_1^{(m)} - u_3^{(m)}),$$
$$p_m = 2, \quad q_m = 3: \ (0,\ 0,\ v_2^{(m)} - v_3^{(m)},\ u_3^{(m)} - u_2^{(m)},\ v_3^{(m)} - v_2^{(m)},\ u_2^{(m)} - u_3^{(m)}).$$

(1) The idea of studying this case was suggested to us by L.A. Bunimovich.

In other words, e_m is obtained from the velocity vector of the system $(u_1^{(m)}, v_1^{(m)}, u_2^{(m)}, v_2^{(m)}, u_3^{(m)}, v_3^{(m)})$ by projection on the two-dimensional subspace $D^{(p_m, q_m)}$ in $\mathbf{R}^6$ and a rotation through 90° in this subspace. The spaces $D^{(p,q)}$ are defined by the direction vectors $r_1^{(p,q)}, r_2^{(p,q)}$ with coordinates $r_1^{(1,2)} = (1, 0, -1, 0, 0, 0)$, $r_2^{(1,2)} = (0, 1, 0, -1, 0, 0)$, $r_1^{(2,3)} = (0, 0, 1, 0, -1, 0)$, $r_2^{(2,3)} = (0, 0, 0, 1, 0, -1)$. We put $D_m = D^{(p_m, q_m)}$. According to (5) the vector $e_m^{(+)} = U_1 \ldots U_m e_m$ lies in $D_m^{(+)} = U_1 \ldots U_m D_m$. If the dimension of $J_+(x)$ from (5) is smaller than 3, then the spaces $D_1^{(+)}, D_2^{(+)}$ are such that in the six-dimensional Euclidean space defined by $(x_1, y_1, x_2, y_2, x_3, y_3)$ there is a two-dimensional subspace L_2 that has a common non-zero vector with each $D_m^{(+)}$, $m \geqslant 1$. This is obviously impossible for spaces $D_m^{(+)}$ $(m = 1, \ldots, k)$ of general position for sufficiently large k.

For the investigation of the specific situation it is sufficient to consider $k = 13$ and run through a finite number of variants of the pairwise collisions of discs with $m = 1, 2, \ldots, 13$. The space D_m takes one of the three values $D^{(1,2)}, D^{(1,3)}, D^{(2,3)}$, while the operator U_m takes one of five values, according to the point A_1–A_5 in whose neighbourhood the m-th collision of the discs occured. The matrices of the operators U_m are computed for the limiting case $r_1 = r_2 = 1/8$, $r_3 = 1/2$, and the results obtained remain true in some neighbourhood of these values of the r_i. The problem is thus reduced to the specific sorting out of the finite number of variants of the pairwise collisions of the discs.

A numerical analysis of all the variants for the system under consideration (it was found that there are several hundred variants) was carried out on a computer, as a result of which 10 variants, for which the position of the spaces $D_m^{(+)}$ $(m = 1, \ldots, 13)$ permit the existence of a two-dimensional space L_2 that has common non-zero vectors with all the $D_m^{(+)}$, were singled out. Furthermore, the set of initial values of the velocity vectors u_i, v_i for which the corresponding space $J_+(x)$ is two-dimensional was found for each of the 10 variants. By (5) $J_+(x)$ coincides with L_2, so $(u_1, v_1, u_2, v_2, u_3, v_3)$ is a two-dimensional subspace V_2 of $\mathbf{R}^6$ defined by the conditions of orthogonality to L_2 and equality to zero of the full momentum of the system. Knowing the numbers of colliding discs for $m = 1, 2, \ldots, 13$, we can exclude certain domains from V_2 by the condition that the scalar product of the vector of relative velocity of the colliding discs and the vector of their centre line is non-zero. This condition is necessary for the realization of the possibility of collision, while the vectors of the centre lines for the limiting values $r_1 = r_2 = 1/8$, $r_3 = 1/2$ are calculated explicitly. After this, non-zero vectors remain in V_2 in only five variants out of the 10. We explicitly analyse these five variants of degenerate trajectories.

1. The velocity vectors have the form $(8t_1, t_2 + 9t_1)$, $(8t_1, -2t_2)$, $(-16t_1, t_2 - 9t_1)$, and the discs collide around A_1 and A_2 (in turn). In this variant all the collisions are close to tangencies, that is, there are almost no

transformations of the velocity vectors, so after a finite number of such collisions there will be a collision around one of the points A_3-A_5, and this will violate the "degeneracy" of this trajectory. Two more variants are the particular cases of this for $t_1 = 0$, when collisions in the neighbourhoods of A_1, A_2, A_4 (in an arbitrary order) are permitted.

2. The velocity vectors have the form $(-4t, 3t)$, $(-4t, 3t)$, $(8t, -6t)$, and the collisions occur around A_1, A_3, A_4 in an arbitrary order. Here all the collisions are also close to tangencies, so the "degeneracy" is violated for the same reasons as in variant 1.

3. The velocity vectors have the form $(24t, 7t)$, $(24t, 7t)$, $(-48t, -14t)$, $t > 0$, the collisions occur around A_1, A_3, A_5, A_1, A_2, A_4 successively, and this cycle is repeated. On the part (A_1, A_2, A_4, A_1) of this cycle the vector of relative velocity of discs 1 and 2 is successively equal to $(-72t, 0)$, $(-36t, -48t)$, and $(0, 0)$, that is, discs 1 and 2 diverge on the parts (A_1, A_2) and (A_2, A_4), while their relative velocity is close to zero on the part (A_4, A_1). Hence a second collision is possible only on condition that the time interval of the motion in (A_1, A_2, A_4) is negligibly small compared to that in (A_4, A_1). But then in the last part the third disc, moving with velocity $(-72t, -21t)$ relative to the first two discs, will considerably distance itself from the first disc, and since the next three collisions in neighbourhoods of A_3, A_5, A_1 occur in a relatively short time interval, discs 1 and 3 will not be brought close to each other, and the motion in the next part (A_1, A_2, A_4) will take a relatively long time. This will lead to an increase in the distance between discs 1 and 2 and to the impossibility of their repeated collision after A_4, since their relative velocity will again be close to zero. Thus the cycle $(A_1, A_3, A_5, A_1, A_2, A_4)$ cannot be repeated with the given initial velocity vectors more than twice, after which the "degeneracy" is inevitably violated.

The above analysis shows that there is a constant $\Lambda > 1$ such that for every point x of the phase space of the system under consideration $\varkappa_{t,0}(x) > \Lambda$ for some $t = t(x) > 0$. Hence $W = \emptyset$, and Conditions A, B in Theorems 5, 5′ are satisfied, that is, these theorems are true for all points $x \in \partial\mathfrak{M}$ of the boundary of the phase space except those whose trajectories $T_1^n x$ hit the manifold of discontinuity R twice for $n = 0, \pm 1, \pm 2, \dots$. But such points form countably many submanifolds of codimension 2 in $\partial\mathfrak{M}$, so the complement to them is linearly connected, and by Theorems 5 and 5′ they form one ergodic component of T_1.

References

[1] D.V. Anosov, Geodesic flows on closed Riemannian manifolds of negative curvature, Trudy Mat. Inst. Steklov. **90** (1967), 1-210. MR **36** # 7157.
= Proc. Steklov. Inst. Math. **90** (1967), 1-235.

[2] ——— and Ya.G. Sinai, Certain smooth ergodic systems, Uspekhi Mat. Nauk 22:5 (1967), 107-172. MR **37** # 370.
= Russian Math. Surveys **22**:5 (1967), 103-167.

[3] V.I. Arnol'd, Small denominators and problems of stability of motion in classical and celestial mechanics, Uspekhi Mat. Nauk **18**:6 (1963), 91-192. MR **30** # 943.
= Russian Math. Surveys **18**:6 (1963), 85-191.

[4] N.N. Bogolyubov and B.I. Khatset, On some mathematical questions of the theory of statistical equilibrium, Dokl. Akad. Nauk SSSR **66**:3 (1949), 321-324. MR **11**-40.

[5] ———, D.Ya. Petrina, and B.I. Khatset, Mathematical description of the equilibrium state of classical systems on the basis of the canonical ensemble formalism, Teoret. Mat. Fiz. **1** (1969), 251-274. MR **56** # 17652.
= Theoret. and Math. Phys. **1** (1969), 194-212.

[6] L. Boltzmann, Lectures on gas theory (Translation from German), University of California Press, Berkeley, CA, 1964.
Translation: *Lektsii po teorii gazov,* Gostekhizdat, Moscow 1956.

[7] M.I. Brin and Ya.B. Pesin, Partially hyperbolic dynamical systems, Izv. Akad. Nauk SSSR Ser. Mat. **38** (1974), 170-212. MR **49** # 8058.
= Math. USSR-Izv. **8** (1974), 177-218.

[8] L.A. Bunimovich and Ya.G. Sinai, The fundamental theorem of the theory of scattering billiards, Mat. Sb. **90** (1973), 415-431. MR **51** # 3395.
= Math. USSR-Sb. **19** (1973), 407-423.

[9] R.L. Dobrushin, Description of a random field by means of conditional probabilities and conditions for its regularity, Teor. Veroyatnost. i Primenen. **13** (1968), 201-229. MR **37** # 6989.
= Theory Probability Appl. **13**:4 (1968), 44-57.

[10] I.P. Kornfel'd, Ya.G. Sinai, and S.V. Fomin, *Ergodicheskaya teoriya,* Nauka, Moscow 1980. MR **83a**:28017.
Translation: Ergodic theory, Springer-Verlag, New York-Heidelberg-Berlin 1982.

[11] N.S. Krylov, *Raboty po obosnovaniyu statisticheskoi fiziki* (Works on the justification of statistical physics), Izdat. Akad. Nauk SSSR, Moscow-Leningrad 1950.

[12] K. Kuratowski, Topology, Academic Press, New York-London 1966. MR **36** # 840.
Translation: *Topologiya,* Mir, Moscow 1966. MR **36** # 839.

[13] Ya.B. Pesin, Characteristic Lyapunov exponents and smooth ergodic theory, Uspekhi Mat. Nauk **32**:4 (1977), 55-112. MR **57** # 6667.
= Russian Math. Surveys **32**:4 (1977), 55-114.

[14] V.A. Rokhlin, Lectures on the entropy theory of measure preserving transformations, Uspekhi Mat. Nauk **22**:5 (1967), 3-56. MR **36** # 349.
= Russian Math. Surveys **22**:5 (1967), 1-52.

[15] ——— and Ya.G. Sinai, The structure and properties of invariant measurable partitions, Dokl. Akad. Nauk SSSR **141** (1961), 1038-1041. MR **27** # 2604.
= Soviet Math. Dokl. **2** (1961), 1611-1614.

[16] D. Ruelle, Statistical mechanics. Rigorous results, Benjamin, New York-Amsterdam 1969. MR **44** # 6279.
Translation: *Statisticheskaya mekhanika. Strogie resultaty,* Mir, Moscow 1971.

[17] Ya.G. Sinai, On the foundations of the ergodic hypothesis for a dynamical system of statistical mechanics, Dokl. Akad. Nauk SSSR **153** (1963), 1261-1264. MR **35** # 5576.
= Soviet Math. Dokl. **4** (1963), 1818-1822.

[18] ———, Classical dynamical systems with countably-multiple Lebesgue spectrum. II, Izv. Akad. Nauk Ser. Mat. **30** (1966), 15-68. MR **33** # 5847.
= Amer. Math. Soc. Transl. (2) **68** (1968), 34-88.

[19] ———, Dynamical systems with elastic reflections. Ergodic properties of scattering billiards, Uspekhi Mat. Nauk **25**:2 (1970), 141-191. MR **43** # 481.
= Russian Math. Surveys **25**:2 (1970), 137-189.

[20] ———, Ergodic properties of a Lorentz gas, Funktsional. Anal. i Prilozhen. **13**:3 (1979), 46-59. MR **81b**:28018.
= Functional Anal. Appl. **13** (1979), 192-202.

[21] ———, Billiard trajectories in a polyhedral angle, Uspekhi Mat. Nauk **33**:1 (1978), 229-230. MR **58** # 7733.
= Russian Math. Surveys **33**:1 (1978), 219-220.

[22] *Sovremennye problemy matematiki. Fundamentalnye napravleniya* (Current mathematical problems. Fundamental directions), vol. 2, VINITI, Moscow 1985.

[23] E. Hopf, Statistik der geodätischen Linien in Manningfaltigkeiten negativer Krummung, Ber. Verh. Sächs. Akad. Wiss. Leipzig **91** (1939), 261-304. MR **1**-243.
= Uspekhi Mat. Nauk **4**:2 (1949), 129-170.

[24] N.I. Chernov, Structure of transversal leaves in multidimensional semidispersing billiards, Funktsional. Anal. i Prilozhen. **16**:4 (1982), 35-46. MR **84m**:58103.
= Functional Anal. Appl. **16** (1982), 270-280.

[25] L.A. Bunimovich and Ya.G. Sinai, Markov partitions for dispersed billiards, Comm. Math. Phys. **73** (1980), 247-280. MR **82e**:58059.

[26] L.A. Bunimovich and Ya.G. Sinai, Statistical properties of Lorentz gas with periodic configuration of scatterers, Comm. Math. Phys. **78** (1981), 479-497. MR **82m**:82007.

[27] H. Federer, Curvature measures, Trans. Amer. Math. Soc. **93** (1959), 418-491. MR 22 # 961.

[28] G. Gallavotti and D. Ornstein, Billiards and Bernoulli schemes, Comm. Math. Phys. **38** (1974), 83-101. MR **50** # 7480.

[29] J. Hadamard, Sur l'itération et les solutions asymptotiques des équations differentielles, Bull. Soc. Math. France **29** (1901), 224-228.

[30] G. Hedlund, The dynamics of geodesic flows, Bull. Amer. Math. Soc. **45** (1939), 241-260.

[31] A. Katok and J.M. Strelcyn, Smooth maps with singularities: invariant manifolds, entropy and billiards, Preprint, Univ. Paris-Nord, 1980.

[32] O.E. Lanford and D. Ruelle, Observables at infinity and states with short range correlations in statistical mechanics, Comm. Math. Phys. **13** (1969), 194-215. MR **41** # 1343.

[33] M. Morse, Symbolic dynamics, Mimeographed notes by R. Oldenberger, Institute for Advanced Study, Princeton, NJ, 1966.

[34] Ya.G. Sinai, Development of Krylov's ideas. Appendix to N.S. Krylov, Works on the foundations of statistical physics, Princeton Univ. Press, Princeton, NJ, 1979, 239-281. MR 82e:82001.

Translated by A. Khait

Institute of Theoretical Physics,
Joint Institute of Nuclear Research,
Dubna

Received by the Editors 1 June 1985
(new version 23 February 1987)