

Local Instability of Orbits in Polygonal and Polyhedral Billiards

G. Galperin, T. Krüger, S. Troubetzkoy

Forschungszentrum BiBoS, Universität Bielefeld, D-33619 Bielefeld, Germany

Received June 5, 1992

Abstract. We classify when local instability of orbits of closeby points can occur for billiards in two dimensional polygons, for billiards inside three dimensional polyhedra and for geodesic flows on surfaces of three dimensional polyhedra. We sharpen a theorem of Boldrighini, Keane and Marchetti. We show that polygonal and polyhedral billiards have zero topological entropy. We also prove that billiards in polygons are positive expansive when restricted to the set of non-periodic points. The methods used are elementary geometry and symbolic dynamics.

1. Introduction

We consider billiards inside polygons and polyhedra. Such billiards have been well studied; good surveys are given in the references [Gu2, GaZ and CGa]. None-the-less, many fundamental questions about these systems remain unanswered: are billiards flows in polygons (polyhedral) ergodic, do they all have periodic points, etc.? In this article we study the topological dynamics of polygonal (polyhedral) billiards. The main tool used is coding of orbits by the sequence of edges (faces) they hit. A similar coding was considered in [K]. In the two dimensional case we prove that if two points code to the same forward sequence then they are both periodic. As a corollary we get a strengthening of a result of Katok [K], polygonal billiards have zero topological entropy. As a second corollary we get that any point is either periodic or the closure of its forward orbit includes a vertex, which gives a full topological classification of an a.e. result in [BKM]. Furthermore we show that the billiard flow, when restricted to the set of non-periodic points is positively expansive. These results hold for geodesic flows on polyhedra as well. In the convex three dimensional case more complicated behavior can arise. Every periodic symbolic sequence corresponds to at least one periodic billiard trajectory, however they can correspond to some quasi-periodic billiard trajectories as well. Nonperiodic symbolic sequences do not necessarily determine a unique billiard trajectory. None-the-less we get as a corollary to our results that polyhedral billiards have zero topological entropy and that any

point either has a periodic symbolic orbit or the closure of its forward orbit intersects an edge.

2. Statement of Theorems

2.1 Polygonal Billiards

Suppose $Q \subset \mathbf{R}^2$ is a simply connected polygon, that is Q is compact and ∂Q consists if a finite union of straight line segments. The unit tangent bundle TQ is defined to be the set of all unit tangent vectors with base points in Q which are pointing into Q . The phase space of the billiard flow $\{\phi_t\}$ is TQ . The billiard trajectory of $x \in TQ$ moves with unit speed along a straight line until it reaches the boundary of Q , then instantaneously changes direction according to the rule “the angle of incidence is equal to the angle of reflection,” and continues along the new line. This motion is determined for all times unless the orbit hits a vertex of Q . For any $q \in Q$ the set of directions which hit a vertex is countable. The natural phase volume μ is an invariant measure and the set of points which have finite trajectories has measure 0.

Let Γ be the boundary of Q and $T\Gamma$ be the set of all unit tangent vectors with base points in Γ which are directed inside Q . Let us denote by f the first return map (Poincaré map) to the set $T\Gamma$. This map and its iterates are defined and smooth everywhere except for the vectors whose billiard orbits hits vertices of Q .

Let $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$ be the partition of the set $T\Gamma$ for which all the vectors with base point in one side of Q form an element of the partition. Here n is the number of sides of Q . When it does not lead to confusion we will also label the sides of Q by the same symbols P_1, P_2, \dots, P_n . Let $\Sigma_n^+ := \{1, 2, \dots, n\}^N$ and σ be the shift transformation. Let $TT_1^+ := \{x \in T\Gamma : \text{the forward orbit of } x \text{ never hits a vertex}\}$. The \mathcal{P} -name $w(x) \in \Sigma_n^+$ of the forward orbit of $x \in TT_1^+$ is defined by the formula: $w(x)_i = j$ iff $f^i(x) \in P_j$. Then the symbolic encoding Σ_Q^+ of the billiard map f is given by $\Sigma_Q^+ := \{w \in \Sigma_n^+ : \exists x \in TT_1^+ \text{ s.t. } w = w(x)\}$. For each $w \in \Sigma_Q^+$ let $X(w) = \{x \in TT_1^+ : w = w(x)\}$. For shortness we denote by $x(w)$ an arbitrary element of the set $X(w)$.

We call a set $S \subset TT_1^+$ a *strip* if all $x \in S$ are parallel vectors whose base points form an interval and all $x \in S$ have the same symbolic description w (Fig. 1). If its base points form an open interval we call it an *open strip*.

Theorem 1. *Suppose Q is an arbitrary polygon and $w \in \Sigma_Q^+$ is a periodic sequence with period n . Then all $x(w)$ have periodic trajectories and the set $X(w)$ forms an open strip. If n is even then all $x(w)$ have period n and if n is odd then there is one point $x(w)$ with period n and the rest of the points in the strip have period $2n$.*

Theorem 2. *For an arbitrary polygon Q and for any $w \in \Sigma_Q^+$ which is not periodic the set $X(w)$ consists of only one point.*

As a straightforward corollary of Theorems 1 and 2 we get:

Corollary 1. *The topological entropy of polygonal billiards is zero.*

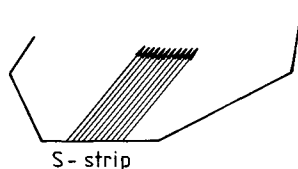


Fig. 1

For a definition of topological entropy in the noncontinuous case see [PP]. It follows that the number of generalized diagonals of length $\leq T$, the number of different words of length n , the number of branches of discontinuity and the number of families of periodic orbits of length n all grow slower than exponentially (“subexponentially” in the terminology of [K]). Katok has shown that the topological entropy of the shift map restricted to the closure of Σ_Q is zero [K].

Let $\pi_1 : TQ \rightarrow Q$ be the natural projection. In [BKM] it was proven that for all $q \in Q$ and for a.e. (with respect to Lebesgue measure) $v \in \mathbf{S}^1$ the closure of the set $\pi_1\{f^i(q, v) : i \geq 0\}$ contains at least one vertex of the polygon Q . As a corollary to Theorem 2 we have the following strengthening of this theorem.

Corollary 2. *For an arbitrary polygon Q and for any $x \in TQ$ either the orbit of x is periodic or the closure of the set $\pi_1\{f^i x : i \geq 0\}$ contains at least one vertex of the polygon.*

The corollary is not true if the second option is replaced by the stronger requirement: the closure of the projection of the forward orbit contains *all* the vertices. A family of counterexamples was constructed in [Ga], where a polygon with n -sides ($n \geq 4$) is constructed for which the closure of the projection of a certain non-periodic orbit contains only two vertices. However the theorem in [BKM] may well still hold under this strengthened requirement.

A flow $\{\phi_t\}$ is said to be *positive expansive* with respect to the metric d if there exists a $\delta > 0$ such that if $x \neq \phi_{t_0}y \ \forall |t_0| < \text{diam}(Q)$ then there exists $t \in \mathbf{R}^+$ such that $d(\phi_t x, \phi_t y) > \delta$. Let $\tilde{T}Q$ be the set of non-periodic points of the billiard flow. Let d be the metric on $\tilde{T}Q$ induced by the Euclidean metric on TQ .

Theorem 3. *The flow $\phi_t|_{\tilde{T}Q}$ is positive expansive.*

The flow is called *spatially positive expansive* if $\phi_t|_{\tilde{T}Q}$ is positive expansive with respect to the Euclidean metric of Q .

Question: For which polygons is the billiard flow spatially expansive?

The billiard flow in any Weyl chamber (i.e. rectangle, $30^\circ - 60^\circ - 90^\circ$ triangle, $45^\circ - 45^\circ - 90^\circ$ triangle, isosceles triangle) is not spatially positive expansive. Two close parallel points will always be close in Q and if their original direction is not rational then they are not periodic.

If Q is an arbitrary polygon of genus $g < \infty$ then we are able to prove slightly less. If w is periodic with period n then the base points of $X(w)$ form at most $gn + 1$ open intervals. If w is not periodic then the set $X(w)$ is a finite set. All points in $X(w)$ are parallel and the base points are separated by at least the minimal diameter (inf of diameters of all orthogonal projections) of the holes of Q . A sufficient condition for the set $X(w)$ consisting of a unique point is the following condition holding for each hole: the maximum of the lengths of the edges of the hole is less than or equal to the minimal diameter of the hole. For example this condition is satisfied by any hole which is a regular polygon. With this extra condition it is not hard to see that if $\text{card } X(w) > 1$ then the set $X(w)$ forms a strip. Corollaries 1 and 2 and Theorem 3 hold in the finite genus case.

2.2 Geodesic Flow on Polyhedra

Suppose that Q is a simply connected polyhedron, that is Q is a simply connected topological surface with a metric which makes it a Euclidean complex (i.e. Q is the

union of Euclidean polygons). We do not necessarily assume that Q is embeddable in \mathbf{R}^3 . If Q_i are the faces of Q then the unit tangent bundle TQ of Q is defined to be the union of the unit tangent bundles TQ_i (in the two dimensional sense as defined above) with the obvious identifications. The set TQ is the phase space of the geodesic flow on the surface Q . As with two dimensional billiards the flow is straight within each face and bounces off boundary edges in the obvious way, that is when the trajectory comes to an edge between two faces Q_i and Q_j then if we “straighten” $Q_i \cup Q_j$ along that edge into a single polygon the geodesic would continue as a straight line from one face to the other. If the geodesic comes to a vertex the motion stops. Again the phase volume μ is invariant and the set of points whose orbits are defined forever has full measure. Again we code the geodesic flow on the surface by the edges it crosses; all definitions are analogous to the billiard inside the polygon and we will not repeat them.

Remark. If we think of a polygon as a degenerate polyhedron then the geodesic flow on the polyhedron which is homeomorphis to a disc is a double cover of the billiard flow in the polygon.

Theorem 4. *Theorems 1–3 and Corollaries 1 and 2 hold for the geodesic flow on the surface of any polyhedron.*

2.3 Polyhedral Billiards

Next we turn to the 3-dimensional case. Suppose Q is a convex simply connected polyhedron embedded in \mathbf{R}^3 . The unit tangent bundle TQ and the billiard flow $\{\phi_t\}$ are defined completely analogously to the polygonal case. In this case the motion is determined for all time if the orbit never hits an edge of Q . Again these points have full measure with respect to the invariant phase volume. For the Poincaré map we introduce symbolic dynamics Σ_Q^+ induced by the partition $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$ of $T\Gamma$ into faces. Here we again abuse terminology and also call the faces P_1, P_2, \dots, P_n . We call a set $S \subset T\Gamma_1$ a *tube* if all $x \in S$ are parallel vectors whose base points form a convex set and all $x \in S$ have the same symbolic description. A tube is called a *polygonal (elliptical) tube* if there is an open polygon (ellipse) F such that $F \subset X(w) \subset \bar{F}$. The rest of the terminology is analogous to the two dimensional case.

Theorem 5. *Suppose Q is an arbitrary convex polyhedron and $w \in \Sigma_Q^+$ is a periodic sequence with period n then the following hold:*

- (a) $\exists x(w)$ s.t. $x(w)$ is periodic with period n ,
- (b) *one of the following two cases hold:*
 - (i) *there exists $q \geq 1$ such that all $y(w) \in X(w) \setminus x(w)$ are periodic with period qn and the set $X(w)$ is an open polygonal tube,*
 - (ii) *the set $X(w)$ is an elliptical tube and there exists a unique point $x(w)$ which is periodic and is at the center of mass of $X(w)$,*
- (c) *if n is odd then only (i) can happen and $q = 2$,*
- (d) *if Q is rational then only (i) can happen.*

Theorem 6. *For an arbitrary convex polyhedron Q and for any $w \in \Sigma_Q^+$ which is not periodic one of the following two statements occurs:*

- (a) *the set $X(w)$ consists of only one point.*
- (b) *the set $X(w)$ is a strip.*

Question: Classify when (b) can occur.

Theorems 5 and 6 yield the following corollary:

Corollary 3. *Convex polyhedral billiards have zero topological entropy.*

The growth rate of the number of generalized diagonals of length $\leq T$, the number of different words of length n , the number of branches of discontinuity and the number of families of periodic orbits of length n again grow slower than exponentially. As another corollary to Theorem 6 we get a three dimensional analogy to [BKM]'s theorem:

Corollary 4. *For an arbitrary convex polyhedron Q and for any $x \in TQ$ either $w(x)$ is periodic or the closure of the $\pi_1\{f^i(x) : i \geq 0\}$ contains at least one point of an edge of Q .*

Question: When does the closure contain a vertex? a whole edge?

The closure cannot contain a vertex if x is an interior point of a strip in case (6b). A second example when the closure does not contain a vertex is a billiard orbit in a cube whose projection on one face is periodic and on a perpendicular face is not periodic.

Next we will construct examples of non-periodic orbits whose closure contains vertices and no other points of an edge. Take Q a nondegenerate (i.e. the angle between adjacent faces is not π) polyhedron in the sense of Sect. 2.2. An enveloping polyhedron Q' has the following properties: each edge P_i of Q lies in some face P'_i of Q' and P'_i is perpendicular to the angular bisector of the two faces of Q which form that edge. Any face P'_i of Q' which is not perpendicularly tangent to an edge P_i of Q does not intersect Q at all. The billiard flow in Q' has the following property: if the orbit starts on Q then its orbit will always stay on Q . Because of this property we call Q a *generalized periodic orbit* in Q' . Any orbit which starts on Q will have the desired property: its closure contains vertices and no other points of an edge.

We give some examples when Theorem (6b) can occur. Suppose Q is a right prism. Then we decompose Q as a direct product along its right angle: $Q = Q_0 \times [0, 1]$. Take a two dimensional billiard in Q_0 . By Theorem 1 any x which is not periodic is the only point with the symbolic description $w(x)$. Notice that since Q is a right prism x has the same symbolic description when considered as a trajectory in Q or in Q_0 . Now in Q consider the billiard trajectory of the vertical strip which is parallel to x . Clearly all the points in this strip have the symbolic description $w(x)$, and other close by parallel points cannot have this symbolic description since in a right prism they would then have the same two dimensional symbolic description as well. One can construct a wider class of examples by attaching arbitrary polygons to the top and bottom of a right prism.

In the convex higher dimensional case the theorems hold with only the obvious modifications. In particular in Theorem (6b) the set $X(w)$ is a co-dimensional 2 convex tube.

3. Unfolding of Trajectories

A simple but important tool which will be used in the proofs of all the theorems is the method of unfolding (or straightening) a trajectory. We will describe this method for polygonal billiards with the changes for polyhedral billiards noted in parenthesis.

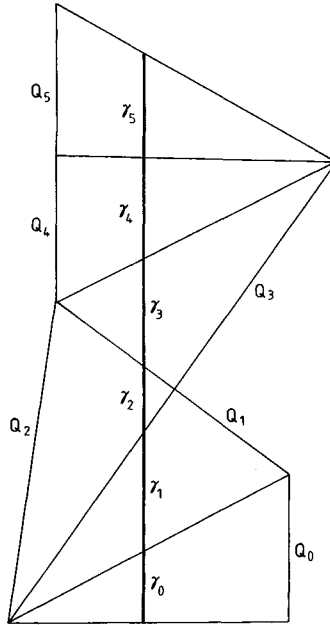


Fig. 2

The method for geodesic flows on polyhedra will be discussed afterwards. Let γ be a billiard trajectory of a point $x \in T\Gamma_1$. A part of γ between two consecutive reflections off the boundary is called a link of γ . Fix the link γ_0 and reflect Q about the side of Γ where γ_0 ends. The reflection of the link γ_1 then becomes the straight continuation of the link γ_0 . Continuing this process for n steps gives a sequence $Q_0 = Q, Q_1, Q_2, \dots, Q_n$ of polygons (polyhedra). The images of the first n links of the trajectory $\gamma_0, \gamma_1, \dots, \gamma_n$ in this unfolding form a straight line segment whose first part is γ_0 . Continuing the unfolding process to infinity gives an infinite sequence of polygons (polyhedra) $Q^\infty(x) := \{Q_0, Q_1, \dots\}$ skewered on a half infinite line whose initial piece is γ_0 (Fig. 2). We call this object an infinite corridor along the (half infinite) line determined by γ_0 . Refolding the corridor the line folds back to the trajectory γ .

For geodesic flows on polyhedra the process is quite similar, but now instead of reflecting the polygon when we reach an edge, we rotate the plane containing the next polygon around this edge until it is in the same plane as the original polygon. This procedure iterated straightens the trajectory into a line. If the polyhedron is embeddable in \mathbf{R}^3 then this can be described in another way: fix a straight line, dip the polyhedron in ink and roll it on the plane so that it always follows the line. The resulting “stamps” will be a straightening of a geodesic on the polyhedron. The following simple lemmas will be used in the proofs of the theorems.

Lemma 1. *If x and y are not parallel vectors than $w(x) \neq w(y)$.*

Proof. Consider the unfoldings of Q for the forward orbits of x and y . The lines in \mathbf{R}^2 (resp. \mathbf{R}^3) determined by x and y may first converge but eventually are linearly diverging. Thus eventually the distance between them is greater than twice the diameter of Q , thus the corridors determined by x and y must be different. The first time they disagree, the reflections must have been made in different edges (resp. faces), thus $w(x) \neq w(y)$. \square

Lemma 2. *If S is a maximal strip (resp. maximal tube in the 3d case) all of whose points have the same symbolic description then the set $f^i S$ never contains a vertex (resp. intersects an edge) for all $i \geq 0$ but $\forall x \in \partial S$ the closure of the set $\{f^i(x) : i \geq 0\}$ contains a vertex (resp. intersects an edge).*

Proof. For $i \geq 0$ the set $f^i S$ never contains a vertex since the points all have the same symbolic description. The maximality of S implies that all boundary points must hit a vertex (resp. edge) or be limit points of a vertex (resp. edge). \square

By definition a strip S is a set of parallel vectors corresponding to a single w whose base points form a segment I , i.e. S can be written as $I \otimes \theta$. In an unfolding the forward orbit of a strip under the flow $\{\phi_t : t \geq 0\}$ will look like an ordinary strip, hence the name. We call this forward orbit S^∞ . As in the two dimensional case a tube S can be written as $C \times \theta$, where C is a convex set. We introduce some convenient notation. Let $l_L(S^\infty)$ and $l_R(S^\infty)$ be the half infinite lines bounding a strip's unfolding. Let $\partial S = \partial C \times \theta$, $\text{diam}(S) = \text{diam}(C)$, etc.

4. Proofs of Two Dimensional Theorems

Proof of Theorem 1. Lemma 1 implies that all the $x(w)$'s are parallel. Suppose w has period n . We assume n is even, if n is odd then it suffices to consider $2n$. Fix $x(w)$ and consider the unfolding $Q^\infty(x)$. Suppose that the sides of Q are numbered so that $x \in P_1$. We claim that $Q_n \in Q^\infty(x)$ has its edge P_1 parallel to the edge P_1 in the original copy $Q_0 \in Q^\infty(x)$. Since n is even the sides Q_0 and Q_n have the same orientation. If they are not parallel then the vectors $f^n x$ and x are not parallel, but since w is periodic this cannot occur because of Lemma 1. Thus Q_n can be gotten from the original copy Q_0 by a translation in the direction of the vector $x(w)$. This implies that all $x(w)$ are fixed points of the mapping f^n . Thus the set $X(w)$ forms a maximal strip. The periodicity of $x(w) \in X(w)$ implies that $x(w)$'s orbit stays a bounded distance away from all vertices and thus the strip must be an open one. Finally if the period of w was odd then Q_n can be gotten from the original copy Q_0 by a translation plus a reflection [C]. Thus the point of reflection symmetry has period n . \square

Proof of Theorem 2. Suppose w is not periodic. We suppose the theorem is not true, that is the cardinality of the set $X(w)$ is strictly greater than one. Lemma 1 implies that all the $x(w)$'s are parallel. Since Q is simply connected all points $z \in T\Gamma$ parallel to this direction whose base points lie between the base points of two members of $X(w)$ also belong to $X(w)$. Thus the set $X(w)$ forms a maximal width strip which we call S . Fix $x(w) \in X(w)$ whose base point lies exactly in the center of S . For emphasis we will write $S(x)$ for S .

If Q is a rational polygon the proof takes an especially simple form and thus we will give it first. Since w is not periodic the trajectory of x must hit some side, say P_1 , infinitely often and since Q is rational it hits P_1 infinitely often with some fixed direction. Call these points $x_i := f^{n_i} x$ and let $S_i(x) := f^{n_i} S(x)$. The strips $S_i(x)$ are of constant width, thus two of them must intersect. Suppose $S_i(x)$ and $S_j(x)$ intersect. They cannot coincide since then the orbit of the $x(w)$ would be periodic. Let $\hat{S}_i(x)$ and $\hat{S}_j(x)$ be the maximal width strips containing $S_i(x)$ and $S_j(x)$ respectively. The strips $\hat{S}_i(x)$ and $\hat{S}_j(x)$ also intersect and since they are parallel part of the boundary

of $\hat{S}_i(x)$ must fall in the interior of the strip $\hat{S}_j(x)$ of vice versa. This contradicts Lemma 2 and finishes the proof in the rational case.

Now we consider the case when Q is not rational. In this case the strip $S(x)$ need not return to any side an infinite number of times each time being parallel to itself. It turns out that we can prove the theorem only using the fact that it returns an infinite number of times being approximately parallel to some direction. A point $x \in T\Gamma_1$ is called *uniformly recurrent* if for any neighborhood $V(x)$ there exists a constant $C > 0$ such that the return times $m_i > 0$ define by $f^{m_i}x \in V$ satisfy $|m_{i+1} - m_i| < C$. Let ω^+ denote the forward limit set. Then the positive width of the strip $S(x)$ implies that the set $Y := \{\omega^+(f^i x)\}$ is closed. Y is also invariant and bounded and thus compact. Consider the dynamical system (Y, f) . Notice that $f|_Y$ is continuous since the set Y is a bounded distance away from the discontinuity points of f . We now apply the following strengthened version of Birkhoff's recurrence theorem which is proven in [F]:

Lemma 3. *If Z is compact and T is continuous then (Z, T) contains a uniformly recurrent point.*

Let x^* be such a point. Assume that the sides of Q are so labeled that $x^* \in P_1$. Then the point x^* is not tangent to the side P_1 since then the forward orbit of x would come arbitrarily close to one of the endpoints of the side P_1 which contradicts the positive width of the strip $S(x)$.

Suppose that the sequence $x_i := f^{n_i}x \rightarrow x^*$. Again we define $S_i(x) := f^{n_i}S(x)$, $\hat{S}_i(x)$ to be the maximal width strip around x_i and $S(x^*)$ to be the maximal width strip around x^* . Consider the image $S^\infty(x^*)$ of the strip $S(x^*)$ in the unfolding $Q^\infty(x^*)$. Fix $\varepsilon > 0$ much smaller than the width of $S(x)$. Consider $Q^\infty(x^*)$ as embedded in \mathbf{R}^2 and let $N_\varepsilon(y)$ be the ε neighborhood of the (base) point y in this embedding. Then let $N_\varepsilon^L = N_\varepsilon^L(S^\infty(x^*)) := N_\varepsilon(l_L(S^\infty(x^*))) \setminus \text{interior}(S^\infty(x^*))$. Define N_ε^R in an analogous fashion. The uniform recurrence of x^* implies that vertices fall inside each of N_ε^L and N_ε^R with bounded gaps between (the heights of) their occurrences. Note that if x^* is periodic then the vertices actually fall periodically on $\partial S^\infty(x^*)$. Consider the intersection of the N_ε 's with the interior of $S_i^\infty(x)$. This intersection contains an ε -wide rectangle whose height, L_i , goes to infinity as $i \rightarrow \infty$. Thus it must eventually include a vertex (Fig. 3). This contradicts Lemma 2. Thus our assumption that the cardinality of the set $X(w)$ is strictly greater than one cannot hold. \square

Proof of Corollary 1. We have shown that the time 0 partition (i.e. the image of \mathcal{P}) is a forward topological generator of Σ_Q . Thus for any ergodic invariant measure ν the metric entropy of the shift map is zero. To apply the variational principle we first take the closure $\bar{\Sigma}_Q$ of Σ_Q . Katok has shown that every ergodic non-atomic shift-invariant measure on $\bar{\Sigma}_Q$ is supported on the set Σ_Q [K]. Therefore applying the variational principle we conclude that the topological entropy of the shift on $\bar{\Sigma}_Q$ is zero (Katok's theorem). Thus the topological entropy of the shift on Σ_Q is also zero. Now $\pi : (T\Gamma_1, f) \rightarrow (\Sigma_Q, \sigma)$ is a continuous factor map; thus if Σ_Q and $T\Gamma_1$ were compact and f were continuous a theorem of Bowen would imply that the topological entropy of the two systems would be the same [B]. However it is not hard to see that Bowen's theorem holds also in this case. We indicate, in the following, the needed modifications to Bowen's proof [B Theorem 17]. The reason that compactness is not needed is that $f|_{\pi^{-1}(y)}$ is an isometry and $\pi^{-1}(y)$ is of bounded diameter for all $y \in \Sigma_Q$. Therefore any ε dense set on $\pi^{-1}(y)$ is an (n, ε) spanning set for all $n \geq 0$ and thus Bowen's sets U_y can be replaced by neighborhoods $B_\varepsilon(\pi^{-1}(y))$ and $\exists \delta > 0$

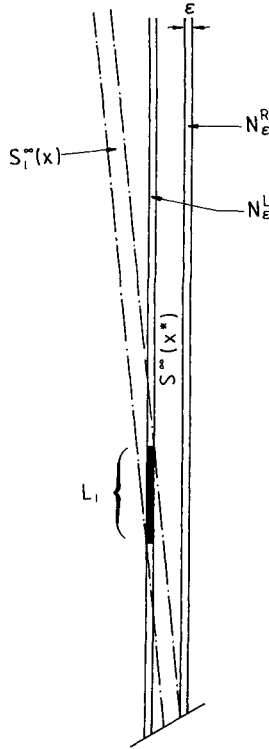


Fig. 3

such that $\pi^{-1}B_\delta(y) \subset U_y \forall y \in \Sigma_Q$. The fact that δ does not depend on y allows us to conclude all of Bowen's arguments which used compactness. \square

Proof of Theorem 3. Suppose $x \neq \phi_{t_0}y$ are not periodic. By Theorem 2, $w(x) \neq w(y)$. Consider the first time in the forward orbit when they hit different sides P_1, P_2 of Q . Call the points of collision x_0 and y_0 . If the orbits of x and y are close enough then we can suppose that P_1 and P_2 share a vertex v_0 . Let α_0 be the angle between P_1 and P_2 . Then the angle between the links ending at x_0 and y_0 or the angle between the links starting at these points is greater than $\alpha_0/2$. Thus the expansivity constant can be taken to be half of the minimal angle between adjacent sides.

Proof of Theorem 4. The proofs of the first two theorems hold verbatim for geodesic flows on polyhedra. For Theorem 3 the expansivity constant must be taken to be half the minimal angle between adjacent faces. \square

5. Proofs of Three Dimensional Theorems

Proof of Theorem 5. (a) Suppose w has period n . Suppose n is even, if n is odd consider $2n$. Consider the unfolding $Q^\infty(w)$. Since n is even the polyhedra Q_n and Q_0 have the same orientation. Since Q is convex and simply connected the set $X(w)$ is convex and bounded. We can extend the mapping f to the boundary of $X(w)$ by continuity from the inside. Since $f^n \overline{X(w)} = \overline{X(w)}$ the Brouwer fixed point theorem tell us that there is a $x(w) \in \overline{X(w)}$ such that $f^n x(w) = x(w)$. This implies that $Q_n \in Q^\infty(w)$ can be gotten from $Q_0 \in Q^\infty(w)$ by a translation along the straightened

trajectory of $x(w)$ plus a rotation (in \mathbf{R}^3) by an angle α around this axis. If $\alpha = 0$ then all $y(w)$ are periodic with period n . If $\alpha \neq 0$ then $x(w)$ must be at the center of mass of $X(w)$ since $X(w)$ is convex and invariant under a rotation around $x(w)$. The point $x(w)$ is periodic with period n . In part (c) we will show that the period of $x(w)$ equals the period of w in the case that the period of w is odd.

(b) (i) If $\alpha = 2\pi p/q$ then $Q_{qn}(w)$ can be gotten from Q_0 by translation alone, and all $y(w)$ are periodic. The shape of the tube $X(w)$ can be described as follows: consider the unfolding of length $qn - 1$ and project all n polygons in this unfolding onto a perpendicular section. This intersection is a polygon. The periodicity of each $y(w)$ implies that its orbit stays a bounded distance from all vertices and thus the tube is open.

(ii) If α is irrational then Q_m is never a translation alone of Q_0 . Thus no point in $X(w) \setminus x(w)$ can be periodic. The shape of the tube is an ellipse swept out by the closest vertex to $x(w)$.

(c) If the period of w is odd then by it is not hard to conclude that $Q_n \in Q^\infty(w)$ can be gotten from $Q_0 \in Q^\infty(w)$ by a translation along the straightened trajectory of $x(w)$ plus a reflection [C]. Thus Q_{2n} is a translation of Q_0 and the claim follows.

(d) If Q is rational then α must be a rational multiple of π . \square

Proof of Theorem 6. Suppose w is not periodic. We suppose the theorem is not true, then the set $S = X(w)$ contains at least three non-colinear points. Lemma 1 implies that all the $x(w)$'s are parallel. The simple connectedness and convexity of Q imply that the base points of the set S form a convex set and S is a tube. Fix $x(w) \in S(x) = S$ whose base point lies in the interior of $S(x)$.

If Q is a rational polyhedron the proof again is simpler and thus we will give it first. Since w is not periodic the trajectory of x must hit some face, say P_1 , infinitely often and since Q is rational it hits P_1 infinitely often with some fixed direction. Call these points $x_i := f^{n_i}x$ and let $S_i(x) := f^{n_i}S(x)$. The tubes $S_i(x)$ are all exactly the same shape, are convex and contain three non-colinear points. Thus two of them must intersect. Suppose $S_i(x)$ and $S_j(x)$ intersect. They cannot coincide since then the orbit of the $x(w)$ would be periodic. Let $\hat{S}_i(x)$ and $\hat{S}_j(x)$ be the maximal tubes containing $S_i(x)$ and $S_j(x)$ respectively. The tubes $\hat{S}_i(x)$ and $\hat{S}_j(x)$ also intersect and thus part of the boundary of $\hat{S}_i(x)$ must fall in the interior of the strip $\hat{S}_j(x)$ or vice versa. This contradicts Lemma 2 and finishes the proof in the rational case.

Now we consider the case when Q is not rational. In this case the tube $S(x)$ need not return to any face an infinite number of times each time being parallel to itself. The minimal diameter of the tube $S(x)$ being positive implies that the set $Y := \{\omega^+(f^i x)\}$ is closed. Y is invariant and bounded and thus compact. Consider the dynamical system (Y, f) . Lemma 3 tells us that there exists a $x^* \in Y$ which is uniformly recurrent. Assume that the faces of Q are so labeled that $x^* \in P_1$. Then the point x^* is not tangent to the face P_1 since then the forward orbit of x would come arbitrarily close to one of the edges of the face P_1 which contradicts the fact that S contains three non-colinear points.

Suppose that the sequence $x_i := f^{n_i}x \rightarrow x^*$. Define $S_i(x) := f^{n_i}S(x)$, $\hat{S}_i(x)$ to be the maximal tube around x_i and $S(x^*)$ to be the maximal tube around x^* . Consider the image $S^\infty(x^*)$ of the tube $S(x^*)$ in the unfolding $Q^\infty(x^*)$. Fix $\varepsilon > 0$ much smaller than the minimal diameter of $S(x)$. As in the two dimensional case if x^* is periodic ε could be taken to be zero. Consider $Q^\infty(x^*)$ as embedded in \mathbf{R}^3 and let $N_\varepsilon(A)$ be the ε neighborhood of the set $A \subset \mathbf{R}^3$. Then let

$N_\varepsilon(S^\infty(x^*)) := N_\varepsilon(\partial S^\infty(x^*)) \setminus \text{interior}(S^\infty(x^*))$. Consider any cross-section C to N_ε which contains the orbit of x^* . C is the union of two strips. The uniform recurrence of x^* implies that C intersects edges with bounded gaps between the heights of their occurrences. Furthermore the bound on the gap size can be taken uniform over all such cross-sections. Consider the intersection of N_ε with the interior of $S_i^\infty(x)$. Some of the above mentioned cross-sections C must contain an ε -wide rectangle in $C \cap S^\infty(x^*) \cap S_i^\infty(x)$ whose height goes to infinity as $i \rightarrow \infty$. Thus some such cross-section must eventually intersect an edge (see proof of Theorem 2). This contradicts Lemma 2. Thus the original strip $S(x)$ could not contain three non-colinear points. \square

Proof of Corollary 3. The proof is the same as of Corollary 1. The time 0 partition is again a forward topological generator of Σ_Q and Katok's proof that no nonatomic measure has support on $\bar{\Sigma}_Q \setminus \Sigma_Q$ works in any dimension without change [K].

Acknowledgements. We would like to thank the Deutsche Forschungsgemeinschaft for their support. We also thank E. Gutkin for posing the problem solved in Theorem 2.

References

- [BKM] Boldrighini, C., Keane, M., Marchetti, F.: Billiards in polygons. *Ann. Prob.* **6**, 532–540 (1978)
- [B] Bowen, R.: Entropy for group endomorphisms and homogeneous spaces. *Trans. Am. Math. Soc.* **153**, 401–414 (1971)
- [CGa] Chernov, N.I., Galperin, G.A.: Billiards and chaos. Moscow: Znania 1991 (in Russian)
- [C] Coxeter, H.M.S.: Introduction to geometry. New York: Wiley 1961
- [F] Furstenberg, H.: Recurrence in ergodic theory and combinatorial number theory. Princeton NJ: Princeton University Press 1981
- [Ga] Galperin, G.A.: Nonperiodic and noteverywhere dense billiard trajectories in convex polygons and polyhedrons. *Commun. Math. Phys.* **91**, 187–211 (1983)
- [GaSV] Galperin, G.A., Stepin, A.M., Vorobetz, Ya.B.: Periodic billiard trajectories in polygons: the mechanism of their appearance. *Russian Math. Surveys* **47**, 5–80 (1992)
- [GaZ] Galperin, G.A., Zemlyakov, A.N.: Mathematical billiards. Moscow: Nauka 1990 (in Russian)
- [Gu1] Gutkin, E.: Billiards on almost integrable polyhedral surfaces. *Erg. Th. Dyn. Sys.* **4**, 569–584 (1984)
- [Gu2] Gutkin, E.: Billiards in polygons. *Physica D* **19**, 311–333 (1986)
- [K] Katok, A.: The Growth rate for the number of singular and periodic orbits for a polygonal billiard. *Commun. Math. Phys.* **111**, 151–160 (1987)
- [PP] Pesin, Ya.B., Pitskel, B.S.: Topological pressure and the variational principle for noncompact sets. *Funct. Anal. Appl.* **18**, 307–318 (1984)

Communicated by Ya. G. Sinai

