ADVANCES IN MATHEMATICS 115, 221–249 (1995)

On the Dual Billiard Problem

S. TABACHNIKOV*

Department of Mathematics, 301 Science Engineering, University of Arkansas, Fayetteville, Arkansas 72701

INTRODUCTION

Let γ be a smooth closed convex curve in the plane with positive curvature and x be a point outside of it. There are two tangent lines to γ through x; choose one of them and reflect x in the point of tangency (Fig. 1). We have defined a map T, which is called the dual billiard map, and the curve γ is called the dual billiard curve.

There are a number of question to ask about this map. Does it have periodic points? Can its orbits have infinity or any points of γ as limit points? For which curves is it integrable? What is the relation of the dual billiard problem to the direct one, i.e., the motion of a point inside a curve according to the laws of geometric optics (Fig. 2)? Can the dual billiard map be defined in a higher dimensional setting? What can be said about the case when γ is a convex polygon rather than a smooth curve? This article addresses some of these questions and contains some (partial) answers to them.

To the best of my knowledge, the dual billiard map appeared for the first time in [D], where the existence of its periodic points was proved. Later, Moser considered the dual billiard map as a crude model for planetary motion [M1, M2]; he pointed out that KAM-theory implies that all of the map's orbits are bounded. The articles [SV, K, GS] contain a sufficient condition for a polygonal γ to guarantee that all orbits of the dual billiard map are bounded. It is weak enough to include regular polygons and polygons whose vertices belong to a lattice.

The contents of this article are as follows. Section I is devoted to the dual billiard problem with a smooth dual billiard curve γ. We start with the observation that area plays a role in the dual billiard problem similar to that of length in the direct one. Namely, periodic trajectories of a dual billiard map are area extrema of circumscribed polygons about γ (as

* E-mail address: serge@uafsysb.uark.edu.

0001-8708/95 $12.00

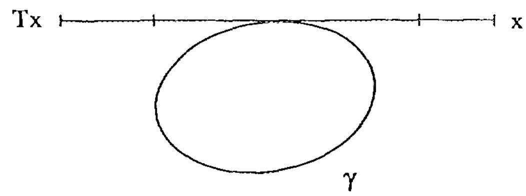

FIGURE 1

periodic trajectories of the billiard ball are length extrema). A Poincaré–Birkhoff-type result on the existence of periodic points of T with arbitrary period $n \geqslant 3$ follows (this is Day's proof [D]).

To explain this "length area duality," we consider both problems on a sphere, where they become projective-dual. The distance between points equals (constant times) the angle between the dual lines, so the perimeter of a polygon equals the sum of angles of the dual one. The latter sum is related to the area of the polygon via the Gauss–Bonnet theorem; hence, periodic trajectories of the dual billiard map are extrema of area. This property holds as the radius of the sphere tends to infinity, though the duality between the two problems breaks in the limit case.

A fundamental property of the dual billiard map is its area-preserving [D, M1, M2]; in view of higher dimensional generalizations we prefer to say that T is a symplectomorphism. We show that T is a small perturbation of an integrable map near infinity (this is the idea of Moser) and near γ as well; it follows from KAM-theory that it has invariant curves arbitrarily close to infinity and to γ. Hence, each orbit is separated both from γ and infinity. In the proof we follow [L], where the existence of invariant curves (caustics) of the (direct) billiard map for a convex billiard curve was established.

On the other hand, the dual billiard problem (as the direct one) is a particular case of a general situation in symplectic geometry considered by

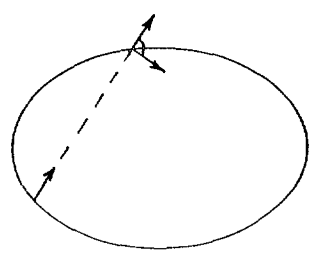

FIGURE 2

Melrose (the so-called hexagonal diagram [Me, Ar1, Ar2]). One of the consequences is the following result:

The asymptotics of areas bounded by simple n-periodic trajectories of T is given, as $n \to \infty$, by

$$A_n \sim a_0 + a_1/n^2 + a_2/n^2 + \cdots + a_i/n^{2i} + \cdots,$$

where a_0 is (of course) the area bounded by γ, and

$$a_1 = \tfrac{1}{24}\left(\int_0^L K^{1/3}\, ds\right)^3,$$

where $K(s)$ is the curvature of γ, s is the length parameter, and L is the length of γ. (The above integral, being affine-invariant, is called the affine length.) In the proof we follow [MM], where the asymptotics of the length spectrum of the billiard map was discovered.

We also include the dual billiard map into a one-parameter family of symplectomorphisms T_α, $\alpha \in [0, \pi/2]$. To construct the map T_α, let x be a point outside of γ and consider a ray, emanating from x, which makes the angle of α with γ. Then $T_\alpha(x)$ is the point of the ray, reflected in γ (as if γ were a mirror), at the same distance from the point of reflection as x (Fig. 3). $T_{\pi/2}$ is the identity and T_0 is the dual billiard

Section II pursues the analogy between the dual billiard problem and the direct one to higher dimensions. A trajectory of a higher dimensional billiard, which is close to the boundary hypersurface $\Gamma^{n-1} \subset \mathbb{R}^n$, is "almost" a geodesic line on it. Geodesic lines are characteristics of the unit-impulse hypersurface of the cotangent space $T^*\Gamma$. This suggests the following generalization of the dual billiard map.

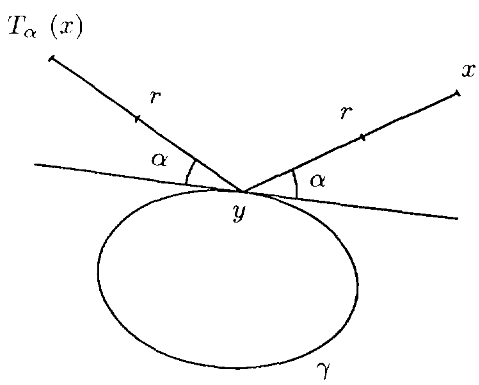

FIGURE 3

Let $M^{2n-1} \subset \mathbb{R}^{2n}$ be a smooth closed strictly convex hypersurface in a linear symplectic space. At each point $y \in M$ the oriented characteristic line is defined to be the skew compliment of the tangent hyperplane $T_y M$. It turns out that for each point x outside of Γ there exists a unique $y \in M$ such that yx has the characteristic direction at y. Define the dual billiard map as the reflection of x in y. This map possesses the same basic property: it is a symplectomorphism. Moreover, it is included into a one-parameter family of symplectomorphisms T_α, $\alpha \in [0, \pi/2]$, analogous to the one constructed in the planar case.

We prove that if p is an odd prime, then T possesses periodic points of period p. The proof makes use of the Morse theory and is close to considerations in [G]. We conjecture that T has periodic orbits in an arbitrary small neighborhood of M, which, as the period tends to infinity, approximate closed characteristics of M (this might be an alternative approach to a proof of the existence of closed characteristics). We also construct the dual billiard map for convex polyhedra in $\mathbb{R}^{2n}$; it consists of reflections in its even-dimensional faces.

Section III is devoted to polygonal dual billiards in the plane. We start with a simple observation that even-periodic points are stable in the sense that they have polygonal open neighborhoods consisting of points with the same dynamics (see also [SV, GS]).

Then we discuss in detail the case of an affine-regular pentagon. We show that the set of periodic points is dense and consists of open regular decagons and pentagons. On the other hand, the set of points with infinite orbits has a nice self-similar ("fractal") structure. It consists of circular "webs," which have the Hausdorff dimension of $6/\ln(\sqrt{5}+2) \approx 1.24$. Each infinite orbit is dense in the corresponding circular web (see the computer picture—Fig. 18).

The dynamics of points with infinite orbits inside the first circular web is the substitution dynamics corresponding to the sequence

$$0010100 \quad 0010100 \quad 000 \quad 0010100 \quad 000 \quad 0010100 \quad 0010100 \cdots,$$

which is invariant under the substitutions

$$0 \mapsto 0010100, \qquad 1 \mapsto 000.$$

This sequence is non-periodic. We find the periods of (stable) periodic points inside the first circular web; they form two sequences:

10, 70, 410, 2470, ...; generic term $= 10(8 \cdot 6^n + (-1)^{n-1})/7$;
stable neighborhood is a regular decagon;

10, 50, 310, 1850, ...; generic term $= 10(6^n + (-1)^{n-1})/7$;
stable neighborhood is a regular pentagon.

Computer experiments suggest that such a self-similarity holds for all regular n-gons (except for $n = 3, 4, 6$; see Fig. 19) and the Hausdorff dimension of the set of non-periodic points is strictly between 1 and 2.

I. The Planar Dual Billiard Problem: A Symplectic Approach

1. We start with the following simple observation [FT]. Let α be a convex plane curve. Fix a positive real c and consider the one-parameter family of chords of α which cut segments of areas equal to c from α. Let γ be the envelope of this family.

LEMMA. *The point of tangency of each chord to γ divides it into equal halves.*

Indeed, all triangles bounded by two infinitely close chords and the curve α have equal areas (Fig. 4); hence, their sides are equal.

Suppose a smooth closed convex curve α is known to be invariant under a dual billiard map. Is it possible to recover the dual billiard curve γ?

COROLLARY. *There exists a one-parameter family of the desired curves γ, namely, the envelopes of the above-considered families of chords of α.*

Note that a γ thus constructed may have double points and cusps, but it still has a well-defined tangent line at each point (that is, is a front) (Fig. 5). A similar method of recovering a billiard curve γ by its caustic α is known [Be]: if γ is the set of points x such that the sum of tangent

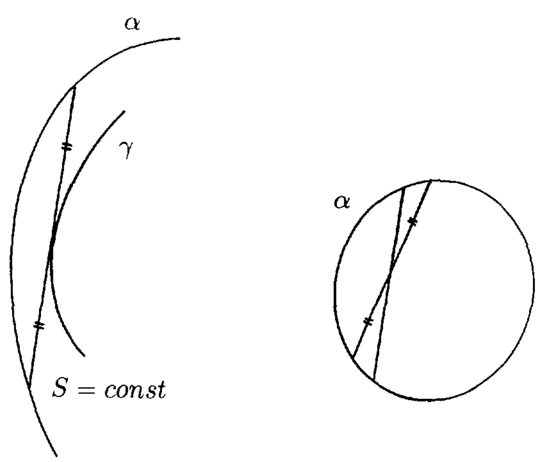

FIGURE 4

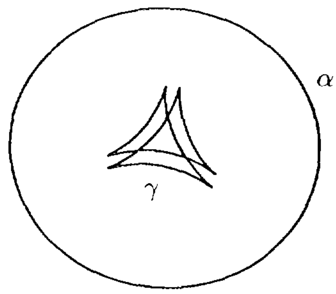

FIGURE 5

segments from x to α and the arc of α between the tangency points is fixed, than a billiard trajectory tangent to α remains tangent to it after reflection in γ (Fig. 6, "the string construction").

2. Let T be the dual billiard map corresponding to a curve γ.

THEOREM [D]. *T has periodic orbits of all periods $n \geqslant 3$.*

Proof. Consider the set of convex circumscribed k-gons with $k \leqslant n$. This set is compact; therefore, the area function possesses its minimum in it. The vertices of this minimal polygon form a T-orbit. If it has less than n vertices, one can decrease its area by cutting off a triangle.

One can also construct periodic orbits of different winding numbers (Fig. 7). The above consideration is absolutely parallel to Birkhoff's well-known proof that a convex billiard has periodic trajectories of all periods $\geqslant 2$ (replace "circumscribed" by "inscribed," and "minimal area" by "maximal length" [Bi]).

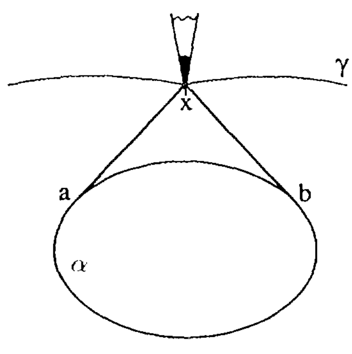

FIG. 6. $xa + ab + bx = \text{const}$.

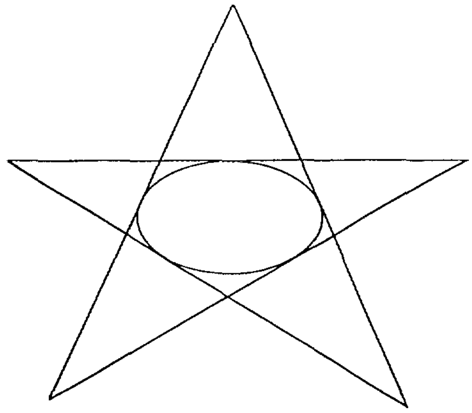

FIGURE 7

Since polygons, joining consecutive images of the dual billiard map, are extrema of the area functional, and since the value of a function is constant on its critical set, we obtain the following result.

COROLLARY. *Suppose there is a one-parameter family of n-periodic points. Then the corresponding polygons have the same areas.*

Compare with [GM1] for the case of the billiard map (replace "area" by "length").

It is well-known that caustics of elliptical billiards are confocal ellipses [Be]. A similar property holds for dual billiards.

LEMMA. *Dual billiards corresponding to conics are completely integrable. Invariant curves are concentric homotetical conics (in the case of a parabola this means parallel parabolas with the same axis).*

We conjecture that ellipses are the only curves for which the dual billiard map is integrable.

3. The previous considerations hint at a certain duality between the dual billiard problem and the direct one. To explain it, consider both maps on the sphere, where points are projective-dual to oriented lines: to a pole corresponds its oriented equator.

LEMMA. *The dual and the direct billiard problems are projective-dual.*

Proof. Let γ be a billiard curve, and γ^* its dual curve, consisting of points dual to tangent lines to γ. The billiard map acts on oriented lines (rays): to a ray a corresponds the reflected ray b (Fig. 8). Let l be the line tangent to γ at the point C, where a hits γ. The dual configuration is seen in the right side of Fig. 8. Since the angles between l and a, and l and b, are equal to the distances between A and L, and B and L, B is the image of A under the dual billiard map corresponding to γ^*.

It follows that since closed billiard trajectories are extrema of the length functional on polygons inscribed into γ, the dual polygons are extrema of the "sum of the angles" functional on circumscribed polygons. By the Gauss–Bonnet theorem, this sum equals, up to a constant, the area of the polygon. However, the dual billiard problem fails to be projective-dual to the direct one on the plane (for instance, it is affine-invariant, which is not the case with the direct problem).

4. The most fruitful approach to the billiard problem is based on the observation that the billiard map T is a symplectomorphism. Namely, T acts on unit vectors with the endpoints on γ, as is shown in Fig. 2. If φ is the angle between γ and a unit vector, and s is the length parameter on γ, then T is known to preserve the symplectic form $\sin\varphi \, d\varphi \wedge ds$ [Bi]. The analogy with the dual billiard map suggests it should preserve a symplectic form as well.

Fix the clockwise orientation of γ; let α be the angular parameter on it (i.e., the angle between a tangent ray and a fixed direction) and r be a nonnegative real. Let $M^2 = S^1 \times \mathbb{R}^1_+$ with coordinates (α, r), and define a symplectic form $\omega = r\, dr \wedge d\alpha$ with the potential $\lambda = r^2\, d\alpha/2$. Let F be the map from M to the domain outside of γ; the point (α, r) is sent to the endpoint of the vector of length r tangent to γ at the point α.

LEMMA. *F is a symplectomorphism of M to the domain outside of γ with its standard area form.*

Define a function $f(\alpha, r)$ on M as the area of the triangle bounded by the two lines tangent to γ through $F(\alpha, r)$ and the arc of γ between the tangency points α and α_1 (Fig. 9). Let r_1 be the length of the second tangent segment. Then the dual billiard map $T_1 = F^{-1}TF\colon M \to M$ sends (α, r) to (α_1, r_1).

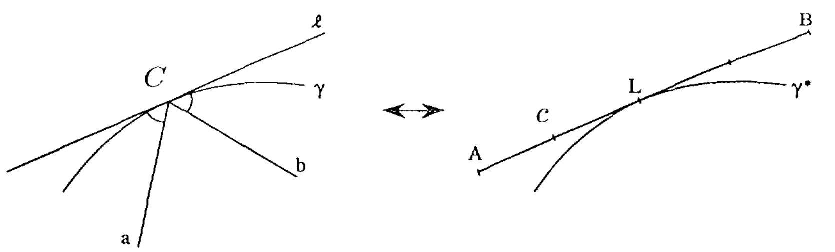

FIGURE 8

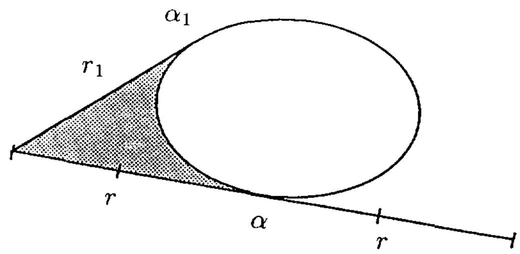

FIGURE 9

THEOREM. *T_1 is an exact symplectomorphism with the generating function f.*

Proof. Observe that

$$\frac{\partial f}{\partial \alpha} = -\frac{r^2}{2}, \qquad \frac{\partial f}{\partial \alpha_1} = \frac{r_1^2}{2}.$$

Hence,

$$T_1^* \lambda - \lambda = \frac{\partial f}{\partial \alpha_1} d\alpha_1 + \frac{\partial f}{\partial \alpha} d\alpha = df.$$

In particular, $T_1^* \omega - \omega = d^2 f = 0$.

COROLLARY [D, M1, M2]. *The dual billiard map is area-preserving.*

Remarks. (i) Let λ be a real number. Let $F_\lambda(x)$ be the point of the line tangent to γ such that the distance from $F_\lambda(x)$ to the tangency point equals λ times the distance from x to the tangency point. Then F_λ is conformal-symplectic with the distortion coefficient equal to λ^2 [D].

(ii) If γ is an arbitrary front, the dual billiard map becomes a symplectic relation. For example, fronts in Fig. 10 define symplectic involutions of the plane outside some compact sets and consist of their fixed points.

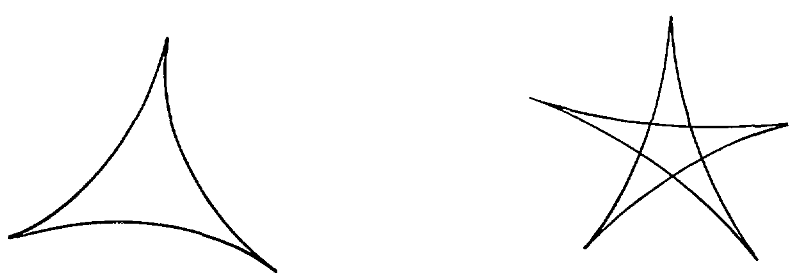

FIGURE 10

5. To prove the existence of invariant curves, we apply the following theorem by Lazutkin [L].

Consider a transformation T of the annulus $S^1 \times \mathbb{R}^1_+$ with coordinates (x modd 1, $y \geqslant 0$), *which has the form*

$$x_1 = x + y + y^{m+1} f(x, y), \quad y_1 = y + y^{m+2} g(x, y), \qquad m \geqslant 1,$$

with smooth function f and g. Assume that for any curve α, homotopic to the curve $y = 0$ and sufficiently close to it, $T(\alpha)$ intersects α. Then T has invariant curves arbitrarily close to $y = 0$.

(This theorem is a KAM-theory-type result.)

First we need a technical result.

LEMMA. *Let T be a transformation of the annulus which in coordinates* (u modd 1, $v \geqslant 0$) *has the form*

$$u_1 = u + \varphi(u)\, v + (v^2), \qquad v_1 = v + \psi(u)\, v^2 + (v^3),$$

where the function φ has no zeros. Then there exist new coordinates

$$x(u, v) = a(u) + (v), \qquad y(u, v) = b(u) v + (v^2),$$

in which T has the form

$$x_1 = x + y + (y^2), \qquad y_1 = y + (y^3).$$

Proof. The equality $x_1 - x = y \bmod (y^2)$ is satisfied iff $a = b/\varphi$. The equality $y_1 = y \bmod (y^3)$ is equivalent to $b\psi + b'\varphi = 0$, which can be solved as $b = \exp(-\int \psi/\varphi)$. Hence, the desired a and b exist.

Now we can prove the existence of invariant curves.

THEOREM. *The dual billiard map T has invariant curves arbitrarily close to γ and arbitrarily far from it.*

Proof. To prove the first claim, we calculate the map T near γ. Let s be the length parameter on γ, and r as before. Let $K(s)$ be the curvature of γ. Then a direct calculation shows that T acts as follows:

$$s_1 = s + 2r - \frac{2}{3}\frac{K'}{K} r^2 + (r^3),$$

$$r_1 = r - \frac{2}{3}\frac{K'}{K} r^2 + \left(\frac{2}{9}\left(\frac{K'}{K}\right)^2 - \frac{2}{3}\left(\frac{K'}{K}\right)'\right) r^3 + (r^4).$$

By the lemma, T has the form

$$x_1 = x + y + (y^2), \qquad y_1 = y + (y^3),$$

and we can apply Lazutkin's theorem, because T is area preserving.

To work "at infinity," let α be the angular parameter and $\rho = 1/r$. Then T "at infinity" can be considered as a map of the annulus (α modd 2π, $\rho \geqslant 0$). We claim that T has the following form:

$$\alpha_1 = \alpha + \pi + f(\alpha)\,\rho + (\rho^2), \qquad \rho_1 = \rho + (\rho^2),$$

where $f(\alpha) < 0$. Indeed, consider the triangle in Fig. 11. Since r is big enough, $\beta \approx l/r = l\rho$, where l is bigger than a positive constant. But $\beta = \alpha + \pi - \alpha_1$, which proves the claim concerning α_1. By the triangle inequality,

$$\rho_1 - \rho = \frac{r - r_1}{r r_1} \leqslant \frac{l}{(r-l)\,r} < \text{const} \cdot \rho^2,$$

which proves the second claim.

It follows that the map T^2 has the form of the lemma, and we obtain its invariant curves in infinity. Finally, if α is invariant under T^2, then $\alpha \cup T(\alpha)$ is invariant under T.

Remark. The formulas show that the dual billiard map near γ is of the form

$$x_1 = x + y + (y^3), \qquad y_1 = y + (y^4),$$

which is also the case in the direct billiard problem [L].

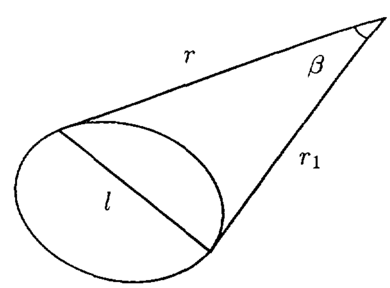

FIGURE 11

COROLLARY. *The orbit of a point is separated both from the dual billiard curve and from infinity.*

6. The dual billiard problem proves to be a particular case of the following general situation, considered by Melrose [Me]. Consider the commutative diagram

$$\begin{tikzcd}
 & X^{2n} & \\
Y^{2n-1} \arrow[ur, hook] \arrow[dd] & & Z^{2n-1} \arrow[ul, hook'] \arrow[dd] \\
 & W^{2n-2} \arrow[ul, hook'] \arrow[ur, hook] \arrow[dl, "p"'] \arrow[dr, "\pi"] & \\
B^{2n-2} & & C^{2n-2} \\
 & \Sigma^{2n-3} \arrow[ul, "\varphi"] \arrow[ur, "\psi"'] &
\end{tikzcd}$$

where X is a sympletic manifold, Y and Z are hypersurfaces, and W is their transversal intersection. The characteristic foliations of Y and Z locally define projections onto the manifolds of characteristics B and C, which inherit sympletic structures from X. The hypersurface $\Sigma \subset W$ consists of points where the restrictions of the sympletic structure of X onto W are degenerate. The characteristic curves of the hypersurfaces $\varphi(\Sigma) \subset B$ and $\psi(\Sigma) \subset C$ are the images of the same curves on Σ, namely, of its characteristics as a submanifold of X [Ar1, Ar2].

The above diagram appears in the direct billiard problem in the following way. Let $\Gamma^{n-1} \subset \mathbb{R}^n$ be a convex billiard, $X = \{(q, p)\}$ be the cotangent bundle $T^*\mathbb{R}^n$ with its standard sympletic structure $dp \wedge dq$, $Y \subset X$ be the unit-momentum hypersurface $p^2 = 1$, and $Z \subset X$ be the hypersurface $q \in \Gamma$. Then B is the space of oriented lines (rays) in $\mathbb{R}^n$ and C is the cotangent bundle $T^*\Gamma$. Characteristics of Y and Z intersect W twice in a neighborhood of Σ, defying two involutions on it, and the billiard map is their composite (see Fig. 2 for the two-dimensional case).

Let $\gamma \subset \mathbb{R}^2$ be the dual billiard curve; $X^4 = T^*\mathbb{R}^2$ with coordinates (x, y, u, v), where (x, y) are coordinates on the plane and (u, v) are momentum coordinates. Define the sympletic form $\omega = dx \wedge dy - du \wedge dv$. Let Y^3 be the unit-momentum hypersurface $u^2 + v^2 = 1$, and let Z^3 consist of points (x, y, u, v) such that the line $l(x, y)$ through (x, y), which is the kernel of $u\,dx + v\,dy$, is tangent to γ (Fig. 12).

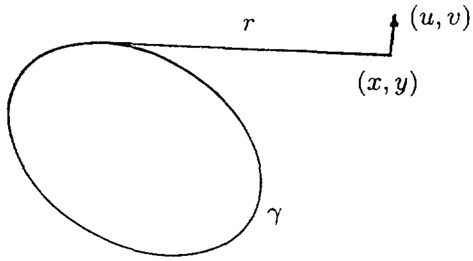

FIGURE 12

LEMMA. *The characteristic of Y through (x_0, y_0, u_0, v_0) consists of unit covectors at (x_0, y_0); the characteristic of Z through (x_0, y_0, u_0, v_0) consists of points (x, y, u, v) with $(x, y) \in l(x_0, y_0)$ and $(u, v) = f(x, y)(u_0, v_0)$ with a certain function f.*

Proof. Let α be the angular parameter in the space of momenta. Since $u^2 + v^2 = 1$ on Y, if follows that $u = \cos\alpha$, $v = \sin\alpha$, and $du \wedge dv = 0$. Therefore, the vector field $\partial/\partial\alpha$ is characteristic for Y.

Let $(x, y) = (\gamma_1(s), \gamma_2(s))$ be the natural parameterization of γ, $p = \sqrt{u^2 + v^2}$, r be the length of the segment tangent from (x, y) to γ (Fig. 12). Then (s, p, r) are coordinates in Z and $u = p\gamma_2'(s)$, $v = -p\gamma_1'(s)$. It follows that

$$\omega|_Z = r\, dr \wedge d\alpha - dp \wedge d\alpha,$$

hence the characteristic field is $\partial/\partial r + r\,\partial/\partial p$. Therefore, s is constant on characteristics.

Include X, Y, Z into the hexagonal diagram. It follows from the lemma that the quotient B^2 is symplectomorphic to $\mathbb{R}^2$ with its area form $dx \wedge dy$, and $C^2 = T^*\gamma$. The surface W^2 consists of the flags

(point outside γ; line tangent to γ through this point).

The intersection of a characteristic of Y with W consists of two lines tangent to γ through a point, and the intersection of a characteristic of Z with X consists of two points of a line tangent to γ at the same distance from the tangency point. Hence, two involutions are defined on W, and the dual billiard map is their composite.

7. Now we are in a position to prove the following result.

THEOREM. *The areas A_n bounded by simple closed broken lines, joining consecutive images of n-periodic points of the dual billiard map, have asymptotics, as $n \to \infty$,*

$$A_n \sim \sum_{i=0}^{l} a_i/n^{2i}.$$

The leading term a_0 is the area bounded by γ; the next term

$$a_1 = \tfrac{1}{24}\left(\int_0^L K(s)^{1/3}\, ds\right)^3,$$

where K is the curvature, s the natural parameter, and L the length of γ.

Proof. Since the dual billiard map T is a composite of the two involutions arising in the Melrose diagram, we apply the analysis undertaken in [MM]. Namely, there exists a smooth nonnegative function h near γ, such that T equals $\exp(h^{1/2}\,\text{sgrad}\, h)$ modulo maps fixing γ to infinite order (sgrad h is the Hamiltonian field of h). Areas A_n are particular cases of the sympletic invariants of [GM2]: if T is an exact symplectomorphism with a generating function f and x is its n-periodic point, then $\sum_{i=0}^{n-1} f(T^i x)$ is a symplectic invariant of T (does not depend on the choice of a potential λ and a function f). The asymptotics as a series in negative even powers of n was established in [MM], where the direct billiard problem was considered.

The changes we have to make are as follows. The coordinates appropriate to the billiard problem are (φ, s) (see Section 4), the sympletic form is $-d\cos\varphi \wedge ds$ and the function h turns out to be $(1-\cos\varphi)\, a(s) + (\varphi^4)$. It was shown in [MM] that the coefficient a_1 of the asymptotics equals, up to a constant,

$$\left(\int_0^L \frac{1}{a(s)}\, ds\right)^3.$$

In the dual billiard problem, we use the coordinates (α, r) of Section 4; the sympletic form equals $d(r^2/2) \wedge d\alpha$. We are looking for the function $h = r^2 a(\alpha) + (r^4)$. The formulas for T show that T has the form

$$\alpha_1 = \alpha + 2rK(\alpha) + (r^2),$$

where $K(\alpha)$ is the curvature as a function of the angular parameter. Since $T = \exp(h^{1/2}\,\text{sgrad}\, h)$—on the level of power series—we conclude that $a(\alpha) = K^{2/3}(\alpha)$. In the natural parameter, $d\alpha = K \cdot ds$, so

$$a_1 = \text{const}\left(\int_0^{2\pi} \frac{1}{a(\alpha)}\, d\alpha\right)^3 = \text{const}\left(\int_0^L K^{1/3}(s)\, ds\right)^3.$$

To find the constant, we consider a circle, where it proves to be $1/24$.

Remark. The length spectrum of the billiard problem is closely related to the spectrum of the Laplace operator with the Dirichlet or Neumann boundary conditions on the billiard curve [GM3]. Is the area spectrum of the dual billiard problem related to the spectrum of any differential operator?

II. Higher Dimensional Dual Billiard Map

1. The motivation for our definition of the dual billiard map in a higher dimensional setting is again the analogy with the direct one. Let $\Gamma^{n-1} \subset \mathbb{R}^n$ be a smooth closed convex hypersurface. Reflection in Γ defines a sympletic billiard transformation of the space of rays N^{2n-2} in $\mathbb{R}^n$ with its natural sympletic structure. To Γ corresponds a hypersurface $\Sigma \subset N$, consisting of rays tangent to Γ. A characteristic of Σ consists of lines tangent to a geodesic line on Γ [Ar1, Ar2]. Let a ray l hit Γ at a point x, and let l_1 be the reflected ray. The rays through x which belong to the plane π, generated by l and l_1, form a line in N. This line is tangent to Σ at $(\pi \cap T_x\Gamma) \in \Sigma$ and has the characteristic direction there.

Translate this to the dual language of the dual billiard problem. Let $M^{2n-1} \subset \mathbb{R}^{2n}$ be a smooth closed strictly convex hypersurface in a linear symplectic space. Given a point x outside of M, find $y \in M$ such that (yx) has the characteristic direction at y, and let $T(x)$ be the reflection of x in y. To justify this definition of the dual billiard map T, we need the following result.

Lemma. *For any x outside of M there exists a unique $y \in M$ as above.*

Proof. Identify $\mathbb{R}^{2n}$ with $\mathbb{C}^n$ and let J be the operator of multiplying by $\sqrt{-1}$ in tangent spaces. If $n(y)$ denotes the unit normal vector at $y \in M$, then the characteristic direction at y is that of $Jn(y)$. Let $\varphi\colon M \times \mathbb{R}^1_t \to \mathbb{R}^{2n}$ send $(y \in M,\ t \geqslant 0)$ to $y + tJn(y)$. We want to show that φ is a homeomorphism onto the exterior of M. Compactify $M \times \mathbb{R}^1_t$ by adding to it the infinite sphere $M \times \{\infty\}$, and compactify $\mathbb{R}^{2n}$ by adding the infinite sphere of all directions. Then φ extends continuously to these spheres as a map of degree 1. Also extend φ to the interior of M as the identity. Thus we obtain a continuous map of closed $2n$-balls, which sends boundary to boundary with degree 1. A standard topological argument shows φ is onto.

To check that φ is one-to-one, assume that the points (y_1, t_1) and (y_2, t_2) have the same image z (Fig. 13). Let π be the plane through y_1, y_2, z and $\gamma = \pi \cap M$. Then γ is convex; $\xi_1 = y_1z$ and $\xi_2 = y_2z$ are vectors tangent to it. The parallel translate of ξ_2 to y_1 has an outward direction

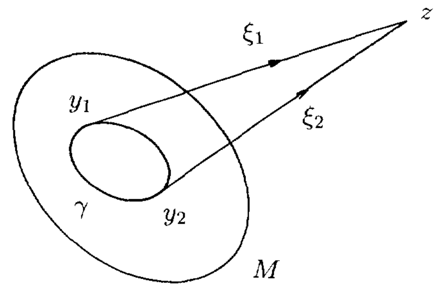

FIGURE 13

with respect to M. By the definition of characteristic directions, the 1-form $i_\xi\omega$ is positive on outward vectors (ω is the linear symplectic form in $\mathbb{R}^n$). Hence, $\omega(\xi_1, \xi_2) > 0$. The same argument implies $\omega(\xi_2, \xi_1) > 0$, which contradicts the skew-symmetricity of ω.

Remark. One can generalize the maps T_α, defined in the Introduction (see Fig. 3), to the higher dimensional situation. To this end, define $\varphi_\alpha: M \times \mathbb{R}^1_t \to \mathbb{R}^{2n}$ to send $(y \in M, t \geqslant 0)$ to $y + t\exp(\sqrt{-1}\,\alpha)\,n(y)$. Then φ_α is a homeomorphism onto the exterior of M, and we set $T_\alpha = \varphi_\alpha \circ \varphi^{-1}_{-\alpha}$.

2. The dual billiard map is defined, and now we shall discuss its symplectic properties.

THEOREM. *The higher dimensional dual billiard map T is a symplectomorphism, and so are all the maps T_α.*

Proof. Consider the product space $\mathbb{R}^{2n} \times \mathbb{R}^{2n}$ with coordinates (x_1, y_1, x_2, y_2), where (x_i, y_i), $i = 1, 2$, are Darboux coordinates in the factors, and equip it with the symplectic structure $\Omega = dx_1 \wedge dy_1 - dx_2 \wedge dy_2$. We shall prove that the graphs of the maps T_α are Lagrange submanifolds of $(\mathbb{R}^{2n} \times \mathbb{R}^{2n}, \Omega)$.

Let $Q \subset \mathbb{R}^{2n} \times \mathbb{R}^{2n}$ be the diagonal with coordinates (q_1, q_2); T^*Q be its cotangent bundle with the momentum coordinates (p_1, p_2); and ω its standard contact structure $dp_1 \wedge dq_1 + dp_2 \wedge dq_2$. The formulas

$$q_1 = \frac{x_1 + x_2}{2}, \quad q_2 = \frac{y_1 + y_2}{2}, \quad p_1 = y_2 - y_1, \quad p_2 = x_1 - x_2$$

define a linear symplectomorphism $\mathbb{R}^{2n} \times \mathbb{R}^{2n} \simeq T^*Q$. The graph of the dual billiard map in T^*Q is contained in the manifold Γ consisting of the pairs

$$(\text{point } y \in M^{2n-1} \subset Q^{2n};\ \text{covector } p \in \operatorname{Ann} T_y M \subset T^*_y Q),$$

i.e., Γ is the conical Lagrange manifold corresponding to $M \subset Q$. Similarly, the graph of T_α is contained in the manifold $\Gamma_\alpha \subset T^*Q$ consisting of the pairs

(the end-point of a normal vector n to M; covector $c\langle n, \cdot \rangle$),

where $c = \cot \alpha$. This is the graph of the differential of the function $f(y) = c \cdot \text{distance}^2(y, M)$, and therefore is a Lagrange manifold.

Remarks. (i) One can define a diffeomorphism F_λ of the exterior of M by extending the characteristic segment through a point to the distance which is equal to λ times its length—compare with Remark (i) in I.4. However, this map fails to be conformal-symplectic.

(ii) If M fails to be convex and even embedded or smooth, but its tangent hyperplane is still well-defined at each point (i.e., M is a front), then the dual billiard map becomes a symplectic relation (just as in the plane case).

(iii) T can also be included into a Melrose hexagonal diagram.

(iv) The above considered manifolds Γ_α are slightly bigger than the graphs of T_α. Namely, Γ_α is the union of the graphs of T_α and T_α^{-1}. This observation is important for the next section.

EXAMPLE. Consider an ellipsoid $\sum (x_i^2/a_i^2 + y_i^2/b_i^2) = 1$; introduce the complex coordinates $z_i = x_i/a_i + \sqrt{-1}(y_i/b_i)$ in $\mathbb{C}^n = \mathbb{R}^{2n}$. A calculation shows that the dual billiard map T acts as follows:

$$z_i \mapsto c_i z_i, \qquad c_i = \frac{a_i^2 b_i^2 - t^2}{a_i^2 b_i^2 + t^2} + \sqrt{-1} \cdot \frac{2t a_i b_i}{a_i^2 b_i^2 + t^2},$$

where the real t satisfies

$$\sum \frac{|z_i|^2}{1 + t^2/a_i^2 b_i^2} = 1.$$

Hence, $|c_i| = 1$ and T is completely integrable with invariant tori $|z_i| = \text{const}$.

This is dual to the well-known fact that a billiard in ellipsoid is integrable, which is the limit case of the integrability of the geodesic flow on ellipsoids [Ar1]. Is there any dual counterpart to the latter?

3. In this section we shall investigate periodic points of the dual billiard map T. We apply a construction introduced in [G]. Consider the space $(\mathbb{R}^{2n})^{(k)} \times (\mathbb{R}^{2n})^{(k)}$, where $X^{(k)}$ means the product of X with itself k times; let

$$(x_1^1, y_1^1, ..., x_1^k, y_1^k, ..., x_2^k, y_2^k)$$

be coordinates in it; let $\Omega = \omega^{(k)} \ominus \omega^{(k)}$ be the symplectic form

$$dx_1^1 \wedge dy_1^1 + \cdots + dx_1^k \wedge dy_1^k - dx_2^1 \wedge dy_2^1 - \cdots - dx_2^k \wedge dy_2^k.$$

Let L be the graph of the cyclic shift, i.e.,

$$L = \{x_2^i = x_1^{i+1}, y_2^i = y_1^{i+1}\}, \qquad i = 1, ..., k,$$

where $k+1$ is understood to be equal to 1. The space $(\mathbb{R}^{2n})^{(k)} \times (R^{2n})^{(k)}$ is symplectomorphic to $T^*Q^{(k)}$, where $Q \subset (\mathbb{R}^{2n}) \times (R^{2n})$ is the diagonal.

LEMMA. *Assume k is odd. Then $L \subset T^*Q^{(k)}$ is the graph of the differential of the function ϕ on $Q^{(k)}$:*

$$\phi(q^1, ..., q^k) = \sum_{i<j} (-1)^{i+j+1} \omega(q^i, q^j),$$

where ω is the linear symplectic structure in $Q \simeq \mathbb{R}^{2n}$.

The proof is a straightforward computation.

Remark. If k is even, the projection of L to $Q^{(k)}$ has a kernel. In this case, one has to apply symplectic reduction (see [G]).

Now we are in a position to establish the existence of periodic points of T.

THEOREM. *If k is an odd prime, then the dual billiard map possesses k-periodic orbits.*

Proof. Let Γ be as in the previous section, i.e., Γ is the union of the graphs of T and T^{-1} (see Remark II.2(iv)). The points of the intersection $\Gamma^{(k)} \cap L$ correspond to k-tuples $(x_1, ..., x_k)$ of points outside of M with $x_{i+1} = T^{\pm 1} x_i$. Since Γ is the conical Lagrange manifold corresponding to $M \subset Q$, the points of $\Gamma^{(k)} \cap L$ are in one-to-one correspondence with the critical points of ϕ restricted to $M^{(k)}$. Since $M^{(k)}$ is compact, this set is non-void.

To show that there exists an "honest" orbit of T, i.e., $(x_1, ..., x_k)$ with $x_{i+1} = Tx_i$, we check that the quadratic differential $d^2\phi$ has a non-zero kernel at each excessive point, i.e., $(x_1, ..., x_k)$ such that $x_i = Tx_{i-1}$, $x_{i+1} = T^{-1} x_i$ for some i. If $\bar{q} = (q^1, ..., q^k) \in M^{(k)}$ is the corresponding critical point of ϕ, then $q^{i-1}=q^i$. Consider the vector ξ tangent to $M^{(k)}$ at $\bar{q}$, all of whose components vanish, except for the $(i-1)$th and ith, and these two are equal to a non-zero vector tangent to M at $q^{i-1}=q^i$. Skew-symmetricity of ω implies then that $d^2\phi(\xi)=0$.

It follows that to maxima (minima) of ϕ there correspond periodic orbits of T, probably multiple ones. Since k is prime, this multiplicity equals one.

Remarks. (i) In the plane case the function ϕ on k-tuples of points of γ differs from the area of the circumscribed polygon through these points.

(ii) It is interesting to study the asymptotic behavior of ϕ as $k \to \infty$.

(iii) We conjecture that T possesses periodic orbits in an arbitrary neighborhood of M. To find such orbits one should restrict ϕ to an open submanifold $N_\varepsilon \subset M^{(k)}$, consisting of k-tuples $(q^1, ..., q^k)$ with $\Sigma\, \mathrm{dist}^2(q_1, q_{i+1}) < \varepsilon$. For ε sufficiently small and k sufficiently large N_ε is a good homotopical approximation of the space ΛM of free loops on M (see [B]).

3. In the plane case, it is obvious how to define the dual billiard map corresponding to a convex polygon rather than a smooth curve. In higher dimensions it is not so obvious. The analogy with the smooth case suggests the following.

Let D be the surface of a convex polyhedron in a linear symplectic space $(\mathbb{R}^{2n}, \omega)$ and let J be the operator of the multiplication of tangent vectors in $\mathbb{R}^{2n}=\mathbb{C}^n$ by $\sqrt{-1}$. Given a point $v \in D$, let C_v be its dual cone, that is, the convex hull of the outward rays normal to the $(2n-1)$-dimensional faces of D containing v. If v belongs to the interior of a k-dimensional face, C_v is $(2n-k)$-dimensional. Apply J to each cone C_v; an argument similar to that of Lemma II.1 shows that the union of JC_v, $v \in D$, covers the domain outside of D.

Let v belong to the interior of a face F. If $\omega|_F$ is degenerate, then $\dim(JF^\perp \cap F) \geqslant 1$. Hence, $\dim(\bigcup_{v\in F} JC_v) < 2n$, and this union is contained in the closure of other cones. Therefore we can ignore such faces, in particular the odd-dimensional faces (if D is in general position, the restriction of ω onto even-dimensional faces is nondegenerate).

An argument, similar to that of Lemma II.1, shows that the interiors of the prisms $\bigcup_{v\in F} JC_v$ with $\mathrm{Ker}\, \omega|_F=0$ are disjoint and cover all the domain outside of D except for a set of dimension $(2n-1)$. We finish with the definition of the dual billiard map T. Given a (generic) point x outside of D, find the (unique) $v \in \mathrm{Int}\, F^{2k} \subset D$ such that $x \in JC_v$ (this cone is $(2n-2k)$-dimensional), and define $T(x)$ to be the reflection of x in v. Study of the dynamics of this map seems to be a challenging problem.

III. Polygonal Dual Billiards

1. In this section we shall study the dynamics of polygonal dual billiards. In this case, the similarity between the dual billiard problem and the direct one breaks completely, and the former exhibits quite new features.

The dual billiard map T is not defined at the points which belong to continuations of the sides of the dual billiard polygon γ. Let X_0 be the set of points for which some iteration of T is not defined. Being a countable union of lines, X_0 is one-dimensional. The rest of the points belong to one of two sets: X_f, the set of points with finite orbits; and X_∞, the set of points with infinite orbits (it is probably appropriate to call them Fatoux and Julia sets). A point x is called stable if it has an open neighborhood which consists of points with the same dynamics as x (i.e., they undergo consecutive reflections in the same vertices of γ as x does). T is a piece-wise isometry, and each component of the complement of X_0 is its domain of continuity. An odd iteration of T on such a domain is a central symmetry; and even one is a parallel translation.

In studying dual billiards the following unfolding method is helpful. Let x be a point outside of γ; choose it as a "system of reference," that is, reflect γ in its vertex which belongs to the left supporting line through x (rather than reflecting the point x) (Fig. 14). This is similar to the well-known unfolding method in the billiard problem.

Lemma (confer [SV, GS]). *Let γ be an n-gon. Even periodic points of T are stable; their stable neighborhoods are open convex k-gons with $k \leqslant 2n$.*

Proof. Let x be a $2N$-periodic point. Then after $2N$ reflections (as in Fig. 14) γ comes to the initial position. The intersection of the exterior

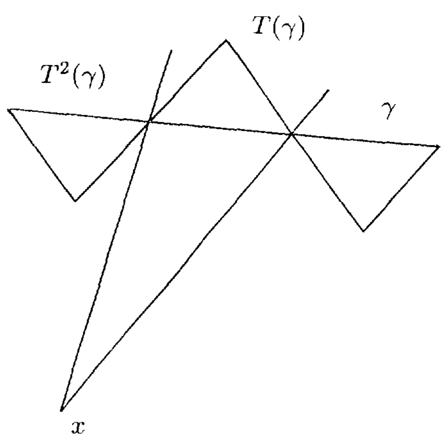

Figure 14

angles of γ, corresponding to each reflection, is the set of points with the same dynamics as x. There are only $2n$ different exterior angles, hence their intersection has at most $2n$ sides.

If x is N-periodic with an odd N, then T^N is the central symmetry of the above neighborhood with the center at x. Then the other points of it are $2N$-periodic.

COROLLARY. *The set X_f is open.*

EXAMPLE [M2, SV]. Let γ be a triangle, a parallelogram, or an affine-regular hexagon. Then the dynamics are trivial: all points have finite orbits.

As was mentioned in the Introduction, [SV, K, GS] contain a sufficient condition for γ to have all orbits of the dual billiard map bounded. In particular, it holds for integer polygons, i.e., polygons whose vertices belong to a lattice. It follows that each orbit is finite for such polygons. The group, generated by reflections in the vertices, is discrete; hence each orbit is discrete and, therefore, finite.

CONJECTURE. For each convex polygon all orbits of the dual billiard map are bounded.

COROLLARY OF THE CONJECTURE. *If a point has an infinite orbit, then it is not stable.*

Proof. Assume U is the maximal connected stable neighborhood. Since T is area-preserving, there exists an even n such that $T^n(U) \cap U \neq \emptyset$. Since the point is non-periodic, T^n is a non-zero shift, and $T^n(U) \neq U$. This contradicts the maximality of U.

2. Now we shall study the case of the (affine-) regular pentagon—the case in which the dual billiard map T exhibits non-trivial dynamics. Consider the computer pictures which show some infinite orbits of T (Fig. 18). We observe a regular global structure and a self-similar local one. We shall make these statements precise.

First, each big, white "circle" (actually, a regular decagon) in Fig. 18 is easily seen to be a periodic stable neighborhood. The map T acts as a cyclic permutation on each "necklace" of these decagons. The domain inside each "necklace" is T-invariant, hence each orbit belongs to an annulus between them.

To explain the global structure of orbits, we use unfolding, i.e., reflect the pentagon γ rather than the initial point 0. Translate the sides of γ to 0, where they define 10 equal angles. Define the coordinates (x, y) in each angle as shown in Fig. 15. We specify the location of γ by the coordinates of its vertex ϑ in which the next reflection occurs, and by an additional parameter $i \in \{0, 1\}$ which distinguishes between two possible orientations of γ (up to parallel translation). As long as ϑ belongs to one angle, reflections occur in the same pair of vertices; the composite of two consecutive ones is a parallel translation over the vector twice the diagonal of γ joining these vertices.

Let (x, y, i) be the coordinates of the vertex ϑ as γ first enters an angle, and let (x_1, y_1, i_1) be its coordinates when γ first enters the next angle. Define the map

$$\phi\colon (x, y, i) \mapsto (x_1, y_1, i_1).$$

Let the side of γ be equal to 1; define the shift S (in each angle) by

$$S\colon (x, y, i) \mapsto (x + 3 + \sqrt{5}, y, i).$$

LEMMA. $S\ \phi = \phi\ S$.

The proof is straightforward.

It follows that the set X_f (the white decagons and pentagons in Fig. 18) and $X_\prime$ (the black "web") are S-invariant (the number $3 + \sqrt{5}$ is exactly the distance between the big, white decagons along a side of the initial pentagon γ). This explains the global regularity of the orbits.

Now we shall study the structure of the set $X_\prime$.

THEOREM. *$X_\prime$ has the Hausdorff dimension of* $\ln 6/\ln(\sqrt{5} + 2) \simeq 1.24$.

Proof. A point x outside of γ reflects in the vertex ϑ if it belongs to the exterior angle at ϑ (Fig. 16). Identify these angles under the rotation about the center of γ through $2\pi/5$. The dual billiard map induces a map—we still

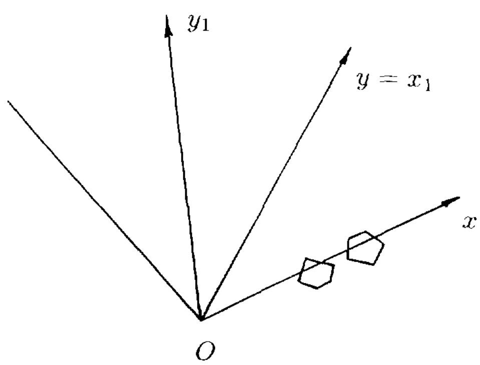

FIGURE 15

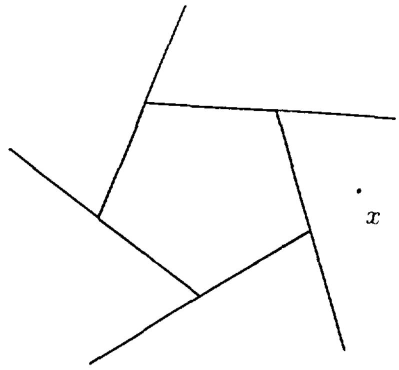

FIGURE 16

call it T—of this quotient angle. In view of the above lemma, it is enough to study T inside the first invariant domain. It consists of the two isosceles triangles OKL and LMN (Fig. 17). The map T acts as the rotation of ΔOKL about the center A through $3\pi/5$, and the rotation of ΔLMN about the center B through $\pi/5$. Let Γ be the dilation with the center O and the coefficient $\lambda = 1/(\sqrt{5}+2)$. It sends B to A, M to M_1, etc.

The big regular decagon in Fig. 17 is invariant under T, and the two big regular pentagons are exchanged by T. Consider the collection of decagons and pentagons, obtained from these ones by applying Γ and T (in all orders). We claim that X_f is the union of their interiors; therefore, X_f and $X_0 \cup X_\infty$ are Γ-invariant.

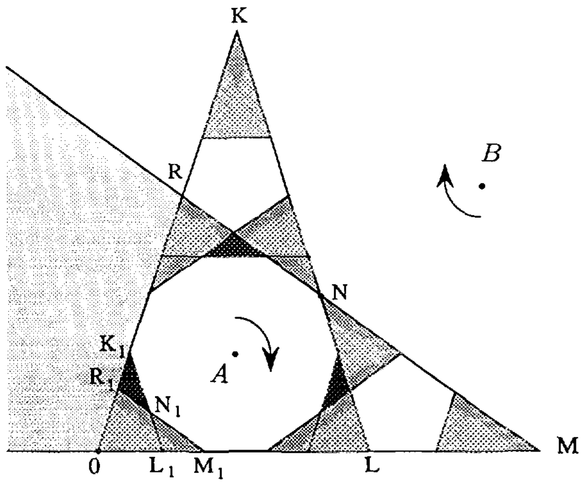

FIGURE 17

Indeed, it is easily seen that T^7 sends ΔOR_1M_1 to ΔOL_1K_1; and T^3 sends $\Delta R_1K_1N_1$ to $\Delta L_1N_1M_1$. Hence, the map $T_1 = \Gamma \circ T \circ \Gamma^{-1}$ equals T^7 on ΔOR_1M_1, and T^3 on $\Delta R_1K_1N_1$. If x is periodic under T, then $x_1 = \Gamma(x)$ is periodic under T_1 (just change the scale); and since T_1 is either T^7 or T^3, x_1 is also periodic under T. Therefore, X_f is Γ-invariant, and so is $X_0 \cup X_{f}$.

Let α be the Hausdorff dimension of X_{f}, which coincides with that of $X_0 \cup X_{f}$, since X_0 is one-dimensional (provided $\alpha \geqslant 1$). Let Y_1 and Y_2 be the parts of $X_0 \cup X_f$ in ΔOKL and ΔLMN, respectively, and u and v be their α-volumes. Figure 17 shows that Y_2 consists of the two parts, λ-similar to Y_1, and Y_1 consists of five parts, λ-similar to Y_1, and three parts, λ-similar to Y_2. Hence,

$$u = 5\lambda^\alpha u + 3\lambda^\alpha v, \qquad v = 2\lambda^\alpha u.$$

It follows that $\lambda^\alpha = 1/6$ and $\alpha = \ln 6/\ln(\sqrt{5}+2)$.

Remark. Computer experiments suggest that for affine-regular n-gons with $n \geqslant 7$, the Hausdorff dimension of X_f is strictly between 1 and 2 (see Fig. 19).

3. Now we are in a position to describe the orbits of T symbolically. Start with decagonal stable domains. Since the map T is the rotation about either A or B, the orbit of a decagon can be encoded by a word in

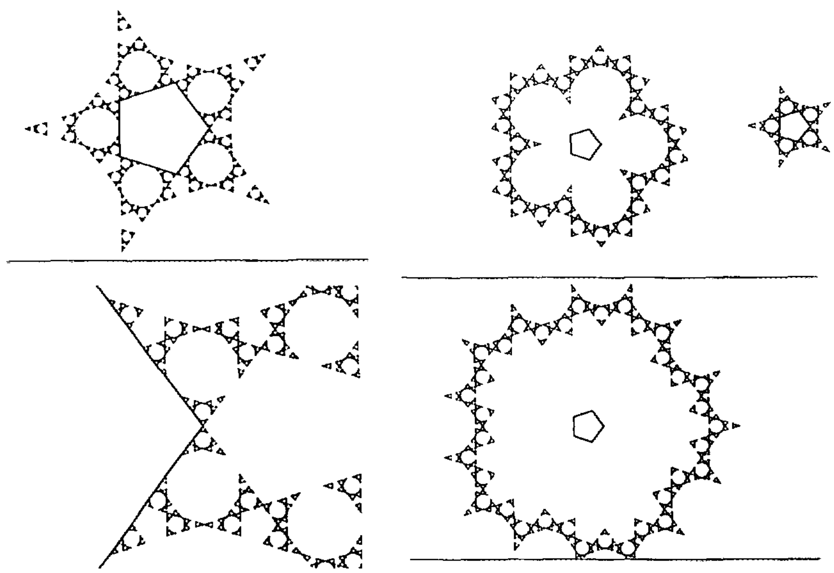

FIGURE 18

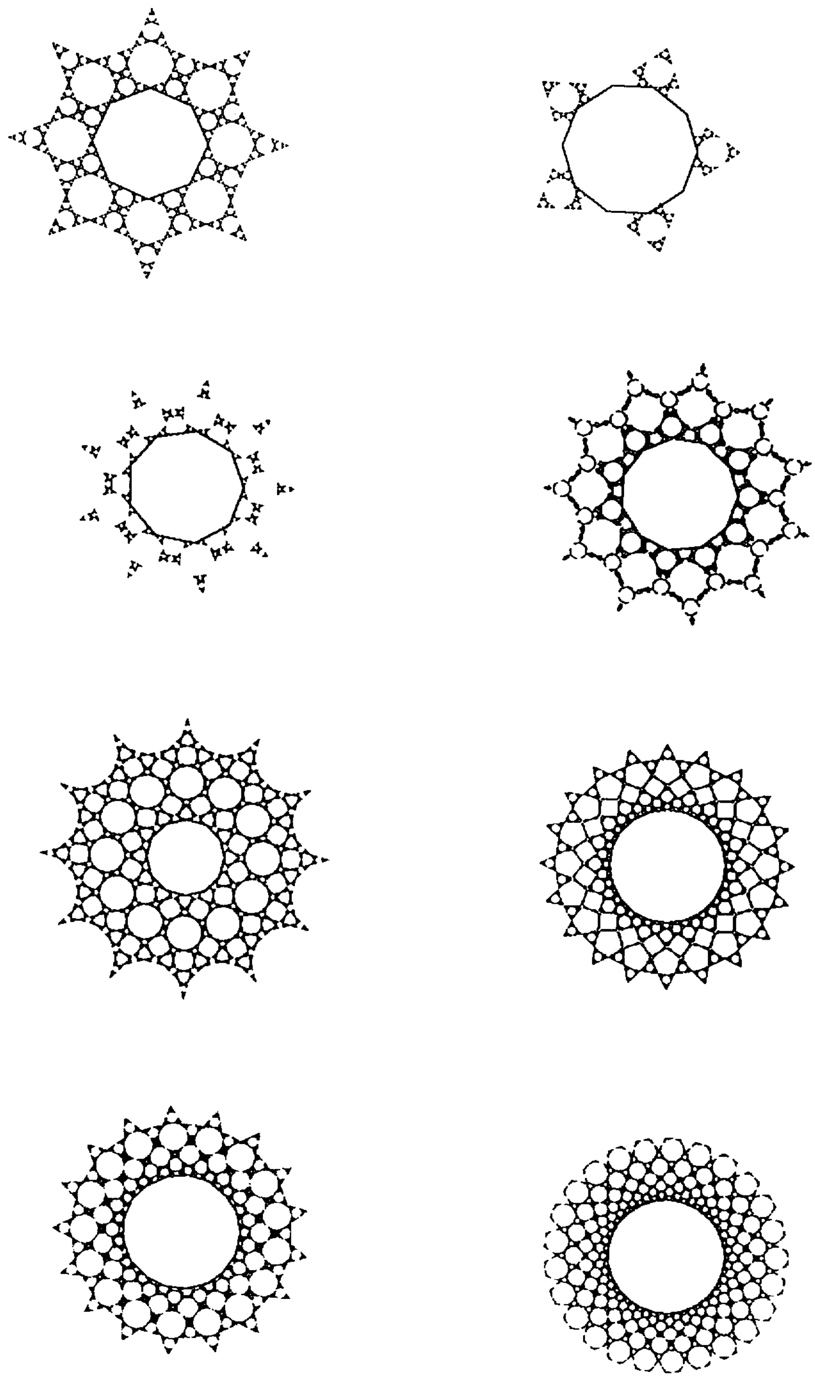

FIGURE 19

the symbols 0 and 1, corresponding to these rotations. The dilation Γ acts as the substitution

$$0 \mapsto 0010100, \qquad 1 \mapsto 000.$$

The symbol 0 corresponds to the biggest decagon (since it is T-invariant). Thus we generate the sequence:

$$0 \xrightarrow{\Gamma} 0010100$$
$$\xrightarrow{\Gamma} 0010100\ 0010100\ 000\ 0010100\ 000\ 0010100\ 0010100 \xrightarrow{\Gamma} \cdots.$$

Let u_n and v_n be the number of 0's and 1's in the nth word, w_n its length.

LEMMA.
$u_n = (6^n + (-1)^{n-1})/7$, $v_n = (2/7)(6^{n-1} + (-1)^n)$, $w_n = (8 \cdot 6^{n-1} + (-1)^n)/7$.

The proof follows from the recurrence

$$u_{n+1} = 5u_n + 3v_n, \qquad v_{n+1} = 2u_n.$$

In the same way, one considers pentagonal stable domains. To them corresponds the sequence

$$01 \xrightarrow{\Gamma} 0010100\ 000 \xrightarrow{\Gamma} \cdots,$$

and the corresponding numbers u'_n, v'_n, w'_n are

$$u'_n = \frac{9 \cdot 6^{n-1} + 2(-1)^n}{7}, \qquad v'_n = \frac{3 \cdot 6^{n-1} + 4(-1)^{n-1}}{7},$$
$$w'_n = \tfrac{2}{7}(6^n + (-1)^{n-1}).$$

Observe that the same recurrences describe the number of decagons and pentagons "of the nth level"; that is, after applying Γ $(n-1)$ times (see the proof of Theorem III.2). It follows that T acts transitively on the set of decagons and the set of pentagons of each level.

To recover periods in the initial dual billiard problem, let us enumerate the exterior angles of the initial pentagon by the elements of $\mathbb{Z}_5$ clockwise. Then the symbol 0 encodes sending a point by the dual billiard map to the next angle, and 1 to the one after the next. Hence, T^{w_n} (or $T^{w'_n}$) corresponds to adding $u_n + 2v_n$ (or $u'_n + 2v'_n$) to the number of the initial exterior angle. Since

$$u_n + 2v_n = (-1)^{n-1} \bmod 5, \qquad u'_n + 2v'_n = 2 \cdot (-1)^n \bmod 5,$$

one has to apply T^{w_n} (or $T^{w'_n}$) five times before a stable decagon (or pentagon) returns to its initial position. In the case of a decagon, T^{5w_n} is a central symmetry, so we have to double the number of iterations. This proves the following result.

THEOREM. *Periods of points with decagonal and pentagonal stable neighborhoods are equal, respectively, to*

$$\tfrac{10}{7}(8\cdot 6^{n-1}+(-1)^n) \qquad \textit{and} \qquad \tfrac{10}{7}(6^n+(-1)^{n-1}).$$

Finally, the following result holds.

THEOREM. *Each infinite orbit is dense in X_∞ (recall that we discuss only the first stable domain of T).*

Proof. Let $x, y \in X_\infty$. We want to show that the orbit of y comes arbitrarily close to x. There exist points x_1 and y_1 of the stable (say) decagonal neighborhoods of the same level, which are sufficiently close to x and y, respectively. Since T acts transitively on decagonal neighborhoods of the same level, $T^k(y_1)$ is sufficiently close to x_1 for some k. If the level of neighborhoods is big enough, $T^k(y)$ is close enough to $T^k(y_1)$, and we are done.

4. The direct limit of the above sequences of the symbols 0 and 1, encoding finite orbits of T, is the infinite sequence

$$\xi = 0010100\ 0010100\ 000\ 0010100\ 000\ 0010100\ 0010100\cdots,$$

which is invariant under the substitution

$$\Gamma\colon 0 \mapsto 0010100, \qquad 1 \mapsto 000.$$

In view of the existence of infinite orbits, the following result is hardly surprising.

THEOREM. *ξ is non-periodic.*

Proof. Assume ξ has a period, and let $\eta(0, 1)$ be the minimal one. Let $\bar{0} = 0010100$, $\bar{1} = 000$. Then $\xi(0, 1) = \xi(\bar{0}, \bar{1})$ (it is the Γ-invariance). Making the preperiod bigger, we assume the period starts with $\bar{0}$ (it should contain at least one $\bar{0}$, because otherwise the rest of ξ would consist of the symbols 0). We claim that $\eta(0, 1)$ is a word in $\bar{0}$ and $\bar{1}$. If not, then $\eta(0, 1)$ ends with an incomplete word $\bar{0}$ or $\bar{1}$:

$$\begin{array}{lcl} \eta = 0010100\cdots 00 & \vdots & 10100 \\ & \vdots & \qquad \nearrow 0010100 \\ & \vdots & \qquad \searrow 000 \\ \eta = 0010100\cdots\ \ 0 & \vdots & 00 \end{array}.$$

It is clear that wherever the period ends, the next segment of ξ cannot start with the words $\bar{0}$.

Hence, $\eta(0, 1) = \eta_1(\bar{0}, \bar{1})$, where η_1 is a shorter word. Then $\eta_1(0, 1)$ is also a period of ξ, which contradicts the minimality of $\eta(0, 1)$.

Remark. The above result and its proof are similar to the well-known case of the Morse sequence

$$0110\ 1001\ 1001\ 0110\ \cdots,$$

invariant under the substitution $0 \mapsto 01$, $1 \mapsto 10$. The shift of this sequence is a particular case of substitution dynamics, i.e., dynamics of shifts of sequences in 0 and 1, invariant under substitutions [Qu]. Our analysis of the dual billiard map for affine-regular pentagons shows that the dynamics of points with infinite orbits is of this type. Does this hold for other regular polygons?

Acknowledgments

I am grateful to numerous mathematicians with whom I discussed dual billiards and learned from, in particular V. Arnold, M. Berger, Ph. Boyland, F. Dekking, D. Fuchs, E. Ghys, A. Givental, B. Grunbaum, E. Gutkin, A. Katok, R. de Llave, J. Moser, C. O'Cinneide, Ya. Sinai, S. Troubetzkoy, A. Veselov, and M. Wojtkowski. My special gratitude goes to I. Monroe, for his beautiful computer pictures.

References

[Ar1] V. Arnold, "Mathematical Methods of Classical Mechanics," Springer-Verlag, New York/Berlin, 1978.

[Ar2] V. Arnold, Contact geometry and wave propagation, Monograph 34 de l'Enseign. Math.

[Be] M. Berger, "Geometry," Springer-Verlag, New York/Berlin, 1987.

[B] R. Bott, Lectures on Morse Theory, Old and New, *Bull. Amer. Math. Soc.* **7** (1982), 331–358.

[Bi] G. Birkhoff, "Dynamical Systems," Amer. Math. Soc. Colloquium Publ., Vol. 9, AMS, Providence, 1927.

[D] M. Day, Polygons circumscribed about closed convex curves, *Trans. Amer. Math. Soc.* **62** (1947), 315–319.

[FT] D. Fuchs and S. Tabachnikov, Segments of equal areas, *Quantum* **2** (1992).

[G] A. Givental, Periodic maps in symplectic topology, *Funct. Anal. Appl.* **23**, No. 4 (1989), 37–52.

[GK] E. Gutkin and A. Katok, Caustics for inner and outer billiards, preprint, 1993.

[GM1] V. Guillemin and R. Melrose, An inverse spectral result for elliptical regions in $\mathbf{R}^2$, *Adv. in Math.* **32** (1979), 128–148.

[GM2] V. Guillemin and R. Melrose, "A Cohomological Invariant of Discrete Dynamical Systems," Christoffel Cent. Vol., pp. 672–679, Birkhäuser, Basel, 1981.

[GM3] V. Guillemin and R. Melrose, The Poisson summation formula for manifolds with boundary, *Adv. in Math.* **32** (1979), 204–232.

[GS] E. Gutkin and N. Simanyi, Dual polygonal billiards and necklace dynamics, *Comm. Math. Phys.* **143** (1991), 431–450.

[K] R. Kolodziej, The antibilliard outside a polygon, *Bull. Polish. Acad. Sci.* **37** (1989), 163–168.

[L] V. Lazutkin, The existence of caustics for a billiard problem in a convex domain, *Math. USSR Izv.* **7** (1973), 185 214.

[M1] J. Moser, Stable and random motions in dynamical systems, *Ann. of Math. Stud.* **77** (1973).

[M2] J. Moser, Is the solar system stable? *Math. Intelligences* **1** (1978), 65–71.

[MM] S. Marvizi and R. Melrose, Spectral invariants of convex planar regions, *J. Differential Geom.* **17** (1982), 475–502.

[Me] R. Melrose, Equivalence of glancing hypersurfaces, *Invent. Math.* **37** (1976), 165–192.

[MT] I. Monroe and S. Tabachnikov, Asymptotic dynamics of the dual billiard map, an example of a semicircle, preprint, 1992.

[Qu] M. Queffelec, "Substitution Dynamical Systems—Spectral Analysis," Lectures Notes in Math., Vol. 1294, Springer-Verlag, New York/Berlin.

[SV] A. Shaidenko and F. Vivaldi, Global stability of a class of discontinuous dual billiards, *Comm. Math. Phys.* **110** (1987), 625–640.

[T1] S. Tabachnikov, Dual billiards, *Russian Math. Surveys* **48**, No. 6 (1993), 75–102.

[T2] S. Tabachnikov, Commuting dual billiards, *Geom. Dedicata* **53** (1994), 57–68.

[T3] S. Tabachnikov, Poncelet's theorem and dual billiards, *L'Enseign. Math.* **39** (1993), 189–194.

[T4] S. Tabachnikov, Billiards: Panoramas et Syntheses, *Soc. Math. France*, No. 1, 1995.