New York Journal of Mathematics

New York J. Math. **3** (1997) 32–47.

Integrable Smooth Planar Billiards and Evolutes

Edoh Y. Amiran

Abstract. Any elliptic region is an example of an integrable domain: the set of tangents to a confocal ellipse or hyperbola remains invariant under reflection across the normal to the boundary. The main result states that when Ω is a strictly convex bounded planar domain with a smooth boundary and is integrable near the boundary, its boundary is necessarily an ellipse. The proof is based on the fact that ellipses satisfy a certain "transitivity property", and that this characterizes ellipses among smooth strictly convex closed planar curves. To establish the transitivity property, KAM theory is used with a perturbation of the integrable billiard map.

Contents

Introduction

By a smooth strictly convex planar domain we shall mean a bounded open region in the plane whose boundary is a C^∞ simple curve with strictly positive curvature. A reflected geodesic in such a domain is a continuous curve consisting of straight line segments in the interior with ends on the boundary at which Snell's law of equal angles with the normal is satisfied. A caustic C for a smooth convex domain is a smooth strictly convex curve homotopic to the boundary with the property that when a (reflected) geodesic is tangent to C each segment of the geodesic is tangent to C. The domain is called integrable near the boundary if some neighborhood of

Received January 7, 1997.
Mathematics Subject Classification. Primary 58F17, secondary 53C22.
Key words and phrases. Billiard ball map, integrable, KAM, caustics, ellipse.

ISSN 1076-9803/97

the boundary is fibered by such caustics. A domain is called integrable if there is a differentiable function on the phase space whose level sets are preserved by the billiard map.

Any elliptic region is an example of an integrable domain: the set of tangents to a confocal ellipse or hyperbola remains invariant under reflection at the boundary. G. D. Birkhoff conjectured that the only integrable convex planar domains are ellipses. An integrable domain with a smooth boundary is also integrable near the boundary, and this article shows that the following theorem applies.

Theorem 1. *When Ω is a strictly convex bounded planar domain with a smooth boundary and is integrable near the boundary, its boundary is necessarily an ellipse.*

The proof is based on the fact that ellipses satisfy a certain "transitivity property", and that this characterizes ellipses among smooth strictly convex closed planar curves [1]. Here we show that a slightly weaker form of this transitivity property is valid when the domain is integrable near the boundary, and that this weaker form suffices for the same conclusion — that the domain is elliptical.

The weaker form of the transitivity property is established by showing that between each caustic, say C, and the boundary there is another caustic, D, so that C is also a caustic for the billiard ball map on D. This uses KAM theory and calculations based on a function used by Lazutkin. (This function is known as H in Mather's variational approach to twist maps.)

In the proof, KAM theory is applied to an auxiliary map which, does not preserve the same symplectic form as the billiard ball map (the map that carries segments of a geodesic to successive segments). Thus we require the stronger form of KAM theory of Moser [15], which uses only the self-intersection condition. Also note that, as a consequence of the theorem, integrability near the boundary translates to integrability in the more traditional sense of the existence of a nontrivial smooth invariant for the billiard ball map.

Anatole Katok suggested that, since certain quantities are conserved (see equation (6.4)) this result might be obtained without KAM theory. Since the perturbation is via a differential equation, the results of Il'yashenko and Yakovenko [7], if extended to the elliptic case, might apply.

The author thanks Richard Melrose for pointing him in the direction leading to this paper and pointing out errors in previous versions. The quality of the content and of the exposition also improved after conversations with John Sylvester, Y. Katznelson, and Eugene Gutkin. The latter patiently asked many useful questions.

1. Invariant Curves and Caustics

Let $\Omega \subset R^2$ be a bounded region with a C^∞ boundary whose curvature is strictly positive. Since $\partial\Omega$ is strictly convex it can be parameterized by tangent angle. Choosing the counterclockwise orientation and the (1,0) direction as the reference for measuring the tangent angle, such regions are determined by their curvature functions – functions $k > 0$ with period 2π and

$$\int_0^{2\pi} k^{-1}(\phi)\cos\phi \; d\phi = \int_0^{2\pi} k^{-1}(\phi)\sin\phi \; d\phi = 0. \tag{1.1}$$

Consider the circle bundle over the boundary, that is, $S_{\partial\Omega}R^2 \subset T_{\partial\Omega}R^2$. The billiard ball map, β, takes (x, u) to (y, v), where x and y lie on the line generated by

u, and v is u translated to y and reflected across the normal to $\partial\Omega$ at y. Identifying the inward pointing portion of $S_{\partial\Omega}R^2$, denoted $S_{\partial\Omega}\Omega$), with $B^*\partial\Omega$ the billiard ball map becomes the boundary map on $B^*\partial\Omega$. It is well known that the boundary map preserves the natural 2-form on $B^*\partial\Omega$. (See, e.g., [5].)

An invariant curve for the billiard ball map on $\partial\Omega$ is a smooth section $\xi : \partial\Omega \to S_{\partial\Omega}\Omega$, which is invariant under the billiard ball map. An invariant curve defines a collection of lines in the plane; for $x \in \partial\Omega$, $(x, v) = \xi(x)$ defines the line through x along v. When $\partial\Omega$ is convex and the invariant curve is near $\partial\Omega$, these lines are the envelope of a convex planar curve whose tangents remain invariant under optic reflection at $\partial\Omega$. (See e.g., [1] for the convexity argument.)

A **caustic** in a domain, Ω, is a planar curve C which is homotopic to the domain's boundary, and whose tangents are invariant under reflection at the boundary. When C is a smooth strictly convex curve inside Ω its tangents define a smooth section of $S_{\partial\Omega}\Omega$. Thus, in a smooth strictly convex planar domain, and in a neighborhood of the boundary, there is a one to one correspondence between invariant curves for the billiard ball map and convex smooth caustics.

Definition 2. The billiard ball map on $\partial\Omega$ is **integrable near the boundary** if there is a neighborhood of $\partial\Omega$ in Ω whose every point belongs to a caustic.

2. The Lazutkin Parameter and the Curvature Relating Operator

For $C \subset \Omega$ a simple and strictly convex curve and a point $p \in \partial\Omega$, there are exactly two points $a, b \in C$ so that the tangent lines to C at a and b go through p. We assume that b follows a according to a fixed orientation of C. We denote the lengths of the line segments between a and p and between b and p by r and l, and the arclength along C between a and b by t (as a consequence of the orientation $t < L(C)/2$ with $L(C)$ the length of C). The **Lazutkin parameter**, Q, of C and $\partial\Omega$ at p (or alternatively at a) is

$$Q(p, \partial\Omega, C) = l + r - t.$$

Lemma 3. [11] *A strictly convex simple closed planar curve $C \subset \Omega$ is a caustic if and only if the Lazutkin parameter of C and $\partial\Omega$ at $p \in \partial\Omega$ is independent of the point p. In fact, with s denoting the arclength along $\partial\Omega$,*

$$\frac{d}{ds}Q(p, \partial\Omega, C) = \cos\theta_+ - \cos\theta_- \tag{2.1}$$

where $\theta_\pm$ are the angles made by the tangent to $\partial\Omega$ at p with the forward and backward rays from C to p (in the notation above, the forward ray is the one from a to p).

The proof uses an appropriate form of Stokes' theorem (see [1]).

Definition 4. For $Q > 0$, and $C \subset \Omega$ a caustic for the billiard ball map on $\partial\Omega$, we say that $\partial\Omega$ is **the Q-evolute of C** if $Q(p, \partial\Omega, C) = Q$ for $p \in C$. We call Q the Lazutkin parameter for C. When p is simply a point in Ω and outside C, the Lazutkin parameter still makes sense and will be denoted $Q(p, C)$.

Let k denote the curvature of C, and v the curvature of $\partial\Omega$, both given in terms of tangent angle (with respect to a fixed direction, say (1,0)). Assume that C is a caustic for the billiard ball map on $\partial\Omega$ with $\partial\Omega$ the Q-evolute of C. Then v can be obtained from k and Q (see §9).

Definition 5. In the setting above the **curvature relating operator** , $L(Q,k)$, is defined by

$$v(\phi) = L(Q,k)(\phi), \quad 0 \le \phi < 2\pi.$$

3. The Weaker Transitivity Property

Assume that C is a smooth strictly convex caustic for the billiard ball map on $\partial\Omega$ with $\partial\Omega \subset R^2$ smooth and strictly convex, and that Ω is integrable near $\partial\Omega$ with a neighborhood of integrability that includes C. Let $\mathfrak{A}$ denote the annulus between C and $\partial\Omega$. We assume that the caustics that fiber $\mathfrak{A}$ are strictly convex. The conclusion of the following theorem describes the transitivity property as it will be used here. Its proof occupies sections four through eight.

Theorem 6. *In the setting of the annulus between C and $\partial\Omega$ as above, let k denote the curvature of C (in terms of tangent angle) and let $R > 0$ be the Lazutkin parameter of C from $\partial\Omega$, that is $L(R,k)$ is the curvature of $\partial\Omega$.*

Then there are sequences $P_n, Q_n \ge 0$ (for $n \ge 0$) with $P_n \to 0$ as $n \to \infty$ and $Q_0 = 0$ ($P_0 = R$), such that

$$L(Q_n, L(P_n,k)) = L(R,k). \tag{3.1}$$

This theorem says that there are closed curves in $\mathfrak{A}$ which are both evolutes of C and caustics. The proof uses KAM theory. In outline, and in our case, the theory gives certain invariant curves for a sufficiently small perturbation of an integrable billiard ball map when the perturbed map, f_ϵ, is a twist map and has the intersection property, that is, for a curve K sufficiently near an invariant curve for the billiard ball map , $f_\epsilon(K) \cap K \neq \emptyset$.

We will define the perturbed map, f_ϵ, by composing a perturbation map, g_ϵ, with the billiard ball map. This perturbation will be chosen so that invariant curves for f_ϵ are both evolutes and caustics.

4. The Perturbed Map

We define the perturbation map with the intersection property in mind, and thus begin by looking at the annulus from the point of view of a fixed caustic C.

In general, a line l in the annulus $\mathfrak{A}$ between C and $\partial\Omega$ corresponds to $(\phi,\theta) \in S_{\partial\Omega}\Omega$ (with $\theta < \pi/2$ – which also determines our choice of ϕ). There is a unique point $q \in l$ where the Lazutkin parameter $Q(\cdot, C)$ has a minimum value. We set $Q(l) = Q(q,C)$ (see Figure 1, below).

In turn, to a point p we associate a line $l(p)$, which is the tangent to the unique caustic (for the billiard ball map on $\partial\Omega$) going through p. Fix a line segment l in $\mathfrak{A}$ and consider p on l. We wish to see what values $Q(l(p))$ can assume.

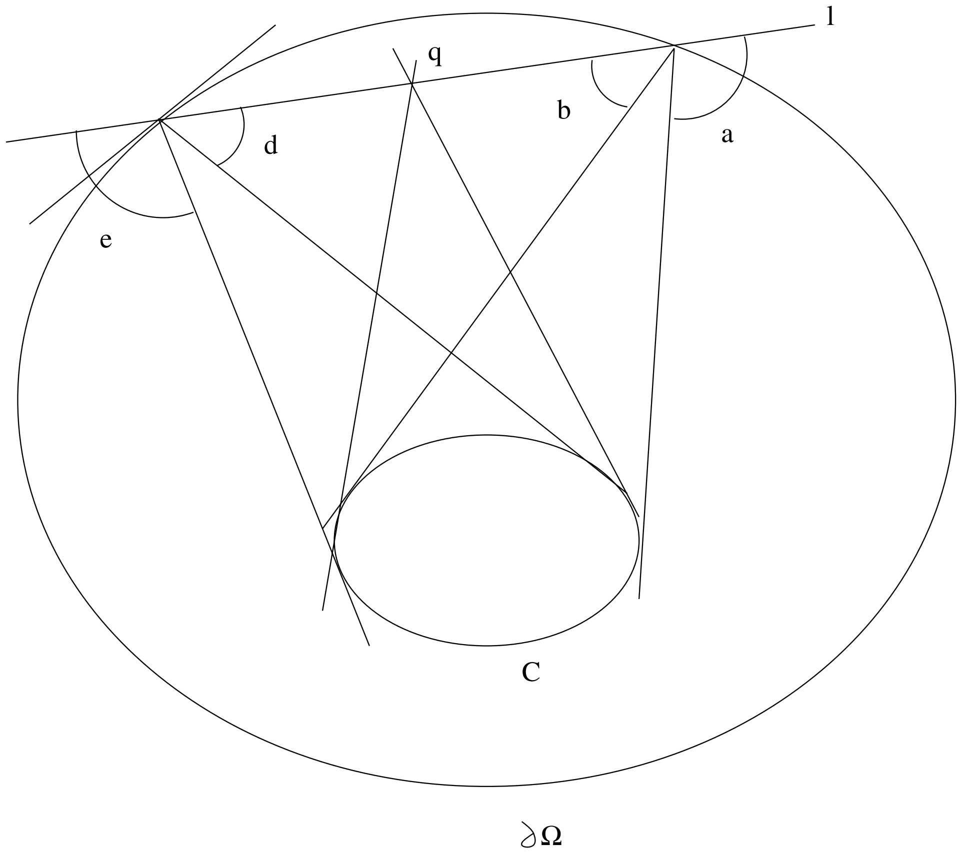

FIGURE 1. $Q(l) = Q(q, C)$ and the caustics D(·)

Since $\partial\Omega$ is a caustic, for $p \in \partial\Omega$, $l(p)$ is the tangent to $\partial\Omega$ at p and since C is a caustic, $Q(l(p)) = Q(p, C)$ which is the largest possible value of Q on the annulus.

We return to the fixed line l, and let q be the point where an evolute, denoted $e(q)$, is tangent to l. Let $p \in l$ be the point where a caustic, denoted $D(p)$, is tangent to l. If $p \neq q$, pick a point $y \in l$ between q and p, and recall that $l(y)$ is tangent to the caustic through y. This line, $l(y)$, crosses the evolute through q. To see this, note that $D(p)$ and $e(q)$ both contain the curve C in their interior and lie on the same side of l, that the tangent to the caustic through y lies outside $D(p)$, and that q lies outside $D(y)$. It follows that if $p \neq q$, and y is as above, $Q(l(y)) < Q(l(q)) = Q(l)$. This is summarized in Proposition 7.

Proposition 7. *In the setting above, when the caustic tangent to l and the evolute tangent to l intersect l at different points, $Q(l(p))$ assumes values both above and below $Q(l)$ as p varies in l.*

REMARKS. 1. The converse to Proposition 7 does not, in general, hold.

2. Proposition 7 shows that the difference between $Q(l)$ and $Q(l(p))$ is a measure of non-agreement between the family of evolutes and the family of caustics. In this

respect, Q is based in the family of evolutes, while $l(p)$ is based in the family of caustics. What is perhaps non-intuitive is that their combination cannot be somehow balanced to hide a difference between these families.

For a given line l and $p \in l$ the point at which a caustic is tangent to l, if the derivative with respect to arclength, s, along l satisfies $\frac{dQ}{ds}(p) \neq 0$, the evolute tangent to l does not intersect l at p (this uses Lemma 3). Hence for some $\epsilon > 0$ we can find a point $y \in l$ with

$$Q(l(y)) \ - \ Q(l) = \epsilon \frac{dQ}{dl}(p). \tag{4.1}$$

In fact, in this case p divides l into two portions, one containing q. The discussion preceding Proposition 7 shows that we can choose y in the portion containing q when $\frac{dQ}{ds}(p) < 0$ and in the portion not containing q when $\frac{dQ}{ds}(p) > 0$.

Let $\epsilon(l)$ be the largest ϵ for which we can solve (4.1) and choose y on the side of p as above. Then for any $0 \leq \epsilon \leq \epsilon(l)$ we can find a solution of (4.1) (i.e., a $y \in l$) which, with the choice above, depends smoothly on ϵ (see the discussion of smoothness in §9). Of course, if $\frac{dQ}{ds}(p) = 0$ then $y = p$ satisfies (4.1) for any ϵ.

Proposition 8. *Assume that Ω is integrable and $\partial\Omega \in C^\infty$ and is strictly convex. Then there is an $r > 0$ such that for $0 \leq \epsilon \leq r$ there is a C^∞ solution, y, of* (4.1).

Proof. Let r be the infimum of the value of $\epsilon(l)$ as l varies over line segments in $S_{\partial\Omega}\Omega$ lying over $\mathfrak{A}$. Because $S_{\partial\Omega}\Omega$ is compact, $r > 0$, and for $0 \leq \epsilon \leq r$ as we vary l (i.e., we vary $(\phi, \theta) \in S_{\partial\Omega}\Omega$ corresponding to l), we can choose solutions of (4.1) which also vary smoothly (in ϕ and θ). More precisely, it follows from the calculations in §8 that when $\partial\Omega$ is C^∞, $Q(p)$ is C^∞ as a function of $p \in \mathfrak{A}$, provided $Q > 0$. Moreover, $y = p$ is a solution of (4.1) for $p \in C$ (and for $p \in \partial\Omega$), and by Proposition 7 and its proof, a solution to (4.1) exists in $\mathfrak{A}$ and approaches $y = p$ as p approaches C. □

For $0 \leq \epsilon \leq r$, we define a map g_ϵ from $S_{\partial\Omega}\Omega$ to itself by choosing for each line l a solution y of (4.1) and setting $g_\epsilon(l) = l(y)$. Our perturbed map is $f_\epsilon = g_\epsilon \circ \beta$.

5. Another View of the Perturbation Map

To better understand the setting for the perturbation of the billiard ball map we examine what occurs in a domain which is not integrable. Let C be a convex caustic for the billiard ball map on $\partial\Omega$. Such caustics exist if $\partial\Omega$ is sufficiently smooth [11], and in fact, one could start with C and generate $\partial\Omega$ as its evolute. Orient C, say counter clockwise. The evolutes of C foliate the annulus $\mathfrak{A}$ between C and $\partial\Omega$, and we can define a map, T, on $\mathfrak{A}$ by $Tx = y$ when y belongs to the same evolute as x and the forward tangent to that evolute at x and the backward tangent to the evolute at y meet $\partial\Omega$ at the same point. The map T is clearly integrable and can be associated with the flow along the tangents to the evolutes.

T can also be defined on $S_{\partial\Omega}\Omega$ by assigning the point x to the line l when the unique evolute tangent to l intersects it at x. The map T is not, in general, in involution with the billiard ball map. If these maps are in involution, then the billiard ball map is integrable and the caustics and evolutes coincide.

The perturbation map of the previous section (where the billiard ball map is integrable, so the family of caustics also foliates the annulus $\mathfrak{A}$) moves points in the annulus in the direction of the flow associated to T, and the "strength" of the perturbation is dependent of the difference between the evolute and the caustic at the point considered, with no perturbation when the tangents to the two families of curves coincide.

The calibration of the perturbation according to the difference between the tangents to the caustics and the tangents to the evolutes is not available in a non-integrable domain. Moreover, were a perturbation map obtainable from T, and had there been an invariant curve for the perturbed map, and were we to show that this invariant curve is an evolute (of C), its relation to $\partial\Omega$ would remain unknown, and so a transitivity property would still not result. In other words, the integrability of the domain is used in defining the perturbation, but also in relating an invariant curve for the perturbed map to the evolutes and caustics.

The "strength" of the perturbation also appears in showing that the image of a curve near a caustic under the perturbed map intersects the curve. A simple argument in the next section shows that the image of a caustic under the perturbed map intersects the caustic. Since the perturbation is "weaker" on a curve in the annulus which is not a caustic, it seems reasonable that the image of such a curve under the perturbed map will also intersect the original curve. Checking this is the main point of the next section.

6. The Intersection Property

For a fixed curve K (a section of $S_{\partial\Omega}\Omega$) which is sufficiently near an invariant curve for the billiard ball map and near $\partial\Omega$, we need to show that $f_\epsilon(K) \cap K \neq \emptyset$. The billiard ball map is area preserving and, in fact, for a curve K differing from an invariant curve, $\beta(K)$ crosses K. This section shows that g_ϵ used for the perturbation cannot disengage this crossing. To do this one needs estimates on the behavior of β near $\partial\Omega$ and on the behavior of g_ϵ. The common currency for these investigations is the parameter Q. The "weakness" of g_ϵ follows from the relationship between Q and the angle of incidence (§9), and between the angle of incidence and the rotation angle (Lemma 10).

Proposition 9. *When I is a caustic, there are at least two points of I invariant under f_ϵ.*

Proof. We say that an evolute E is larger than an evolute F if E contains F in its interior. There are, then, a largest evolute intersecting I and a smallest evolute intersecting I. These intersections are tangential with tangents, say, l and m. Now l and m are fixed by g_ϵ, and since I is a caustic, their preimages under β are in I. □

The rest of this section shows that for a curve K which is not a caustic there is an intersection of $f_\epsilon(K)$ and K and it is, in fact, transversal. We begin with the perturbation map. Let $p \in l$ denote the point where a caustic is tangent to l, and let α denote the signed angle from l to the (tangent to) the evolute through p (see Figure 2.). Then

$$Q(g_\epsilon(l)) - Q(l) = \epsilon|\nabla Q(p)|\sin(\alpha) \tag{6.1}$$

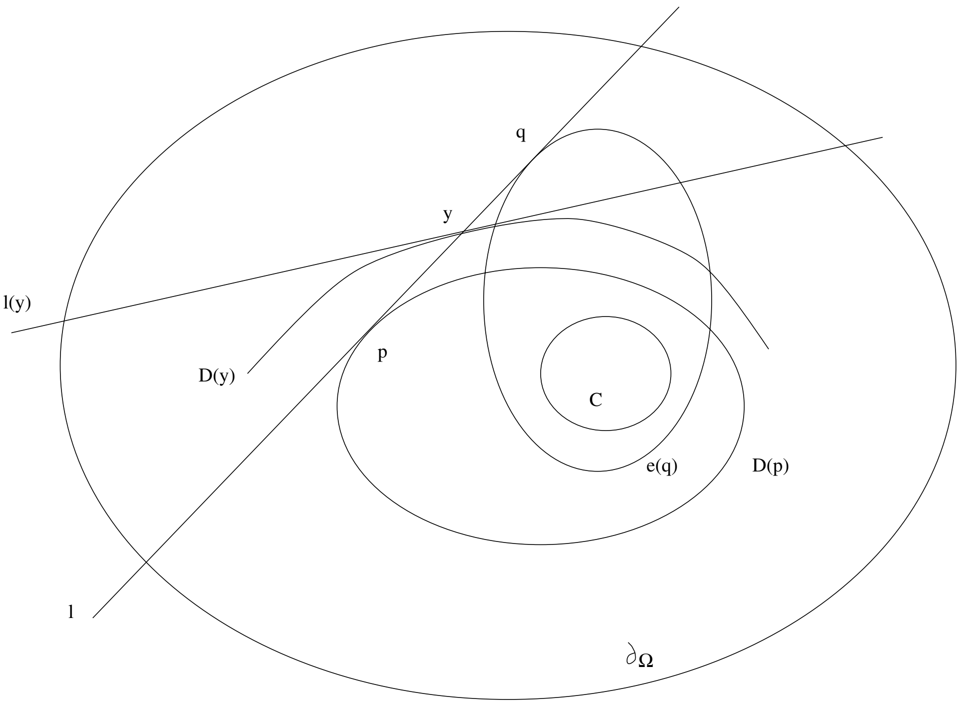

FIGURE 2. Behavior near p

with the gradient taken for Q as a function on R^2.

Suppose that R is the length of the line segment tangent to C at a, L is the length of the line segment tangent to C at b, the radii of curvature of C at a and b are A and B, and the arclength along C between a and b is s. Let Δ be the angle of R (and L) with the evolute through p, and let t parameterize the perpendicular to that evolute which points away from C. Then

$$|\nabla Q(p)| = \frac{dQ}{dt} = \frac{dR}{dt} + \frac{dL}{dt} - \frac{ds}{dt} = \sin(\Delta) + \sin(\Delta) - \big(\frac{A}{R} + \frac{B}{L}\big)\cos(\Delta).$$

When, for two lines at $\phi \in \partial\Omega$, $Q(l) > Q(m)$, the Δ corresponding to l is larger than that corresponding to m, and thus we have

$$\frac{Q(g_\epsilon(l)) - Q(l)}{\sin(\alpha(l))} > \frac{Q(g_\epsilon(m)) - Q(m)}{\sin(\alpha(m))}. \tag{6.2}$$

The distance (in the plane) between $p(l) \in l$ and $g_\epsilon(l)$ is $\epsilon \sin(\alpha) + O(\epsilon^2)$, and the change in the ϕ coordinate between l and $g_\epsilon(l)$ is

$$\Delta\phi \leq \epsilon\nu \cdot k(\phi)\sin(\alpha) + O(\epsilon^2)\ . \tag{6.3}$$

In the above, k is the curvature of $\partial\Omega$, and ν is a positive upper bound for $|\nabla Q|$ in a neighborhood of $\partial\Omega$ with the neighborhood excluding C (to be chosen later). That such a bound exists follows from the differentiability of Q away from C. (See

§9 for a calculation of Q and its properties. The distance along a ray transversal to an evolute varies differentiably with Q at any point away from the fixed curve C.)

The behavior of the billiard ball map can be described in terms of Q by equation (6.5) below.

Let K be a curve in $S_{\mathfrak{A}}\Omega$, and also its representation as a curve in Ω, the latter being convex when K is sufficiently near a caustic, since the caustics correspond to convex curves. Let $q(l)$ be the point of l where an evolute is tangent to l and set $\overline{K} = \{q(l) | l \in K\}$.

Consider the submanifold $\mathfrak{F}$ of $S_{\partial\Omega}\Omega$ generated by the forward flowout from $\overline{K}$ along the lines $l \in K$, and the submanifold $\mathfrak{B}$ generated by the backward flowout from $\overline{\beta(K)}$ along the lines $l_1 \in \beta(K)$. Fix two points on $\partial\Omega$, say $a < b$ in terms of arclength, and consider the manifold M formed by the portions of $\mathfrak{F}$ and $\mathfrak{B}$ which intersect $\partial\Omega$ between these two points. Since Q is a function on $\mathfrak{A}$, dQ is a function on M.

Integrating dQ over the portion of $\mathfrak{F}$ corresponding to $l \in K$ we get $Q(\partial\Omega) - Q(l)$ and integrating dQ over the portion of $\mathfrak{B}$ corresponding to $l_1 \in \beta(K)$ we get $Q(l_1) - Q(\partial\Omega)$. And on the portions of $\mathfrak{F}$ and $\mathfrak{B}$ over $\overline{K}$ and $\overline{\beta}(K)$, dQ is zero.

Let $v(s)$ (resp. $v_1(s)$) be the element of $\mathfrak{F}$ (resp. $\mathfrak{B}$) at $s \in \partial\Omega$. If we denote the angle which v makes with the tangent to $\partial\Omega$ by θ, then the corresponding angle for v_1 is $-\theta$.

Integrating dQ over M, Stokes' theorem yields

(6.4)

$$Q(l(b)) - Q(l(a)) + \int_a^b dQ(v(s))\frac{d\theta}{ds}ds = Q(l_1(b)) - Q(l_1(a)) - \int_a^b dQ(v(s))\frac{d\theta}{ds}ds,$$

where we have used $ddQ = 0$ and, in the second integral, that v_1 is the reflection of v. Differentiating (6.4) we obtain

$$\frac{d(Q|_K)}{ds} + dQ(v(s))\frac{d\theta}{ds} = \frac{d(Q|_{\beta(K)})}{ds} - dQ(v(s))\frac{d\theta}{ds}. \tag{6.5}$$

Here $(Q|_K)$ denotes the function Q restricted to K.

In terms of the ϕ coordinate, for a caustic I, $\beta|_I$ is a monotone rotation, with the average rotation being $2\pi\rho$, where ρ is its rotation number. The following lemma establishes estimates, which are generally known, for the relation of the angle of incidence and the rotation number.

Lemma 10. *Assume that a planar domain Ω has boundary $\partial\Omega$ which is C^3 and has strictly positive curvature. Let θ be the incident angle of a geodesic in Ω whose rotation number ρ exists. Then $\rho = \theta/\pi + O(\theta^2)$.*

Proof. To find the image (s_1, θ_1) of (s, θ) under the billiard ball map we must solve

$$\int_s^{s_1} \cos(\phi(t))dt = l\cos(\phi + \theta) \quad \textit{and} \quad \int_s^{s_1} \sin(\phi(t))dt = l\sin(\phi + \theta) \tag{6.6}$$

for s_1 and l. Then $\phi_1 = \phi(s_1)$ and $\theta_1 = \phi_1 - \phi - \theta$.

Denote the radius of curvature of $\partial\Omega$ by $\chi(s)$. Equation (6.6) gives $\phi_1 \sim \phi + O(\theta)$ so $\phi_1 = \phi + b\theta + O(\theta^2)$, with b a constant. Replacing $\phi(t)$ in the integral in equation (6.6) by its Taylor series at s and eliminating l we get

$$\tan(\phi+\theta) = \frac{\left[b\chi\sin\phi\cdot\theta + \frac{1}{2}b^2\left(\dot\chi\sin\phi + \chi\cos\phi - \dot\chi\sin\phi + \frac{\chi\sin^2\phi}{\cos\phi}\right)\theta^2 + O(\theta^3)\right]}{b\chi\cos\phi\cdot\theta}$$
$$= \tan\phi + \frac{1}{\cos^2\phi}\theta + O(\theta^2).$$

This yields $b = 2$ (independently of s). In addition, $\phi_1 - \phi - 2\theta = u(\phi)\theta^2$, with $u(\phi)$ depending differentiably on ϕ. Consequently (since $\partial\Omega$ is compact), $\rho = \theta/\pi + O(\theta^2)$. □

The rotation number is well defined for geodesics tangent to an invariant curve so it follows from lemma 10 that when a caustic I is sufficiently near $\partial\Omega$, specifically when $\theta < \max(1, (\rho/4)\cdot\max u(\phi))$,

$$0 < \frac{3}{4}\rho\phi(\beta(v)) - \phi(v) < \frac{5}{4}\rho \quad \forall v \in I\ . \tag{6.7}$$

Lemma 10 also yields $\rho = \theta/\pi + \bar u(\phi)\cdot\theta^2$ (where $\bar u(\phi)$ is the average taken over the orbit starting at ϕ). Differentiating this with respect to arclength s, we find that

$$\frac{d\theta}{ds} = -\frac{du}{ds}\frac{\theta^2}{1+2\bar u\theta}.$$

It follows that $|d\theta/ds|$ can be bounded arbitrarily near 0 by choosing a caustic I with a sufficiently small rotation number (i.e., sufficiently near $\partial\Omega$).

Also, when the caustic is near $\partial\Omega$, $dQ(v(s))$ is bounded by $2d(Q|_I)/ds$. Here Q is the Lazutkin parameter from the curve C and $v(s)$ is the direction of the tangent to the caustic, so the directional derivative is evaluated at the corresponding point, s, in $\partial\Omega$. This holds since $v(s)$ approaches the unit tangent direction at $\partial\Omega$ as shown in §9. (See Lemma 16 there, and note that the first term in the curvature relating operator is the identity.)

We return to a general curve K which will be in a neighborhood of a caustic.

Choose a caustic I_0 sufficiently near $\partial\Omega$ so that (6.7) is satisfied and $|dQ(v(s))| < 2|d(Q|_I)/ds|$, and so that $|d\theta/ds| < 1/10$. There is a C^1 neighborhood, $\mathfrak{N}$, of I_0 so that for $v \in K \subset \mathfrak{N}$

$$\frac{1}{2}\rho < \frac{\phi(\beta(v)) - \phi(v)}{2\pi} < \frac{3}{2}\rho\ . \tag{6.8}$$

Finally, let c be the maximum of the curvature of $\partial\Omega$ (which is positive by assumption), and take any $\epsilon < \rho/(2\nu c)$. (Here ν is an upper bound for $|\nabla Q|$ as in (6.3). This choice of ϵ prevents g_ϵ from disengaging K and $\beta(K)$ below by changing ϕ alone.)

Let $(\phi, Q(\phi))$ describe $K \subset \mathfrak{N}$ and $(\phi, Q_1(\phi))$ describe $\beta(K)$. Assume that K is not a caustic (though it must be near I_0).

Since β is area preserving, $K_1 = \beta(K)$ crosses K at least twice. Thus there are tangent angles $a, b \in \partial\Omega$ with $Q_1(a) = Q(a)$, $Q_1(b) = Q(b)$, and $Q_1(\phi) \leq Q(\phi)$ for $a \leq \phi \leq b$.

Let I_1 be the caustic through the endpoint (a or b) where Q is largest. Note that $I_1 \subset \mathfrak{N}$ (in fact, the distance between I_1 and I_0 is smaller than the distance between K and I_0 in terms of the family of caustics).

If we further assume that $f_\epsilon(K) \cap K = \emptyset$, then g_ϵ increases the Q coordinate between a and b for curves near K_1. This follows from (6.3), (6.8) and our choice of ϵ. Note that we may assume α in (6.2) and (6.3) to be non-zero for otherwise I_1 would be tangent to an evolute, and there would be a point in $f_\epsilon(K) \cap K$ as in Proposition 9. In view of equation (6.5) and our estimates on $d\theta/ds$ and $dQ(v(s))$, g_ϵ increases the Q coordinate for K_1 in a *neighborhood* of $[a, b]$.

Since on $[a, b]$ the Q coordinate of I_1 exceeds the Q coordinate of K_1 (I_1 lies "above" K_1), (6.2) applies, and $g_\epsilon(I_1)$ and I_1 do not intersect in a neighborhood of $[a, b]$ and so $f_\epsilon(I_1)$ and I_1 do not intersect in a neighborhood of $[a, b]$. But as $\epsilon \to 0$, g_ϵ approaches the identity, and $f_\epsilon(K) \cap K$ contains points on K in $[a, b]$ because $\beta(K)$ and K cross there. So now we see that, since I_1 is near (I_0 which is near) K, $f_\epsilon(I_1)$ and I_1 *do* intersect in a neighborhood of $[a, b]$. The points of $f_\epsilon(I_1) \cap I_1$ described by Proposition 9, do not, however, change with ϵ. Thus we have arrived at a contradiction, which proves the following lemma.

Lemma 11. *Let f_ϵ be the perturbed map of §4. If* I *is a caustic sufficiently near $\partial\Omega$,* K *is a curve sufficiently close to* I, *and ϵ is sufficiently small, then*

$$f_\epsilon(K) \cap K \neq \emptyset.$$

7. Invariant Curves

Claim 12. *Assume that Ω is a bounded planar domain, integrable near $\partial\Omega$, with $\partial\Omega$ C^∞ and strictly convex. Assume that C is a caustic and the caustics between C and $\partial\Omega$ are C^∞ and strictly convex. Let D denote a caustic which is also an evolute of C. For sufficiently small ϵ, f_ϵ has an invariant curve, K, strictly inside $S_D(intD) \hookrightarrow S_{\partial\Omega}\Omega$ so that f_ϵ restricted to K has an irrational rotation number.*

We will use KAM theory to find the required invariant curve for f_ϵ. Since f_ϵ is a smooth perturbation of the billiard ball map, the theorem needed here – a simplified version of a theorem of Moser [15] – can be stated as follows:

Let $\mathfrak{B}$ be an annulus in R^2 with boundary components the closed oriented curves C and D, and let $h\colon \mathfrak{B} \to \mathfrak{B}$ be a twist map: h is smooth, h(D)=D, h(C)=C, and the rotation numbers of h on C and D, b and a (respectively), satisfy $a < b$.

Assume that an invariant curve for h passes through every point of $\mathfrak{B}$, that is, there are coordinates $0 \leq x \leq 1, 0 \leq y \leq 1$ on $\mathfrak{B}$ with $h(x, y) = (x + y, y)$. (We call y_0 the rotation number on the invariant curve $\{y = y_0\}$.)

Let $h_\epsilon\colon \mathfrak{B} \to \mathfrak{B}$ be a smooth map which is also C^∞ in $\epsilon \in (0, r)$ with $h_0 = h$, and assume that for ϵ sufficiently small h_ϵ is a twist map on $\mathfrak{B}$. Assume also that for any closed curve K sufficiently near an invariant curve for h

$$(h_\epsilon)(K) \cap K \neq \emptyset. \tag{7.1}$$

Then for ϵ sufficiently small, h_ϵ has an invariant curve in $\mathfrak{B}$ with rotation number ω for each ω with $a+\epsilon < \omega < b-\epsilon$ and which is poorly approximated by rationals in the sense that

$$(7.2) \qquad |\omega - 2\pi\frac{m}{n}| > \epsilon n^{-5/2} \quad n, m \in Z, \quad n > 0.$$

NOTE. that as a consequence of the proof of this theorem – see [15] pp. 14 and 15 — (7.1) need only hold in a neighborhood of the invariant curve for h with rotation number ω_0 (satisfying (7.2)) for the conclusion to hold for an invariant curve with rotation number ω_0.

We show that f_ϵ satisfies the hypotheses of the theorem above as a perturbation of the map β.

Proof of Claim 12. Let D and C be as in the hypothesis of Claim 12 and let $\mathfrak{B} \subset S_D(intD)$ be the annulus with boundary components the inclusions of D and C in $S_D(intD)$ (also called D and C). Since C and D are both evolutes of C and caustics, for any ϵ, $f_\epsilon|_C = \beta|_C$, and $f_\epsilon|_D = \beta|_D$, so that $f_\epsilon(C) = C$ and $f_\epsilon(D) = D$ with the rotation numbers of β, showing that for sufficiently small ϵ the map f_ϵ is a twist map of the annulus.

The map $p \mapsto l'$ with l' the tangent to the caustic $D(p)$ through p is smooth since the billiard ball map on $\partial\Omega$ is integrable and the caustic through p is smooth and strictly convex. In fact, since the caustics are strictly convex the collection of lines l' (associated to caustics) can be parameterized by the caustics to which they are tangent and the point of the caustic at which they are tangent (see Lazutkin's [11] which shows that the caustics can be parameterized smoothly by their rotation numbers). Each of the two angles made with l by the tangents to C through a point $x \in l$ is continuous and monotone in x (one increases while the other decreases), showing — by the implicit function theorem — that the map taking l to $q \in l$ used in §4 is smooth. Also, by Proposition 8, we can find a solution of (4.1) which is smooth (as the line l varies). Thus the billiard ball map is perturbed by a smooth map in the annulus.

From §6 (Lemma 11) we know that for ϵ sufficiently small the perturbed map has the intersection property for curves sufficiently near D and sufficiently near a fixed caustic in $\mathfrak{B}$.

Thus, the KAM theory applies, proving Claim 12 for some curve with a rotation number ω satisfying (7.2), in particular, the rotation number is irrational. □

8. Properties of the Invariant Curves

We show that invariant curves for f_ϵ are both evolutes of C and caustics.

Proposition 13. *A caustic, D, in the annulus $\mathfrak{A}$ whose image in $S_{\partial\Omega}\Omega$ is invariant under f_ϵ is an evolute of C.*

Proof. Being a caustic, D is invariant under the billiard ball map, so D is invariant under f_ϵ only if it is invariant under g_ϵ.

For a point $p \in D$, the tangent line at p to D is taken by g_ϵ to the line l' tangent to a caustic $D(q)$ through a point $q \in l$. But since the caustics are convex and do not intersect, either $q = p$ or q lies outside D and $D(q)$ lies outside D. Since D is invariant under f_ϵ it follows that $q = p$.

By Lemma 3, $q = p$ only if the angles between l and the two tangents to C which go through p are equal. Since l is the tangent to D (at p), D is an evolute of C. □

Proposition 14. *A closed invariant curve for f_ϵ on which f_ϵ has an irrational rotation number is (the image of) both an evolute of C and a caustic.*

Proof. Denote by K the image in Ω of the invariant curve for f_ϵ (as well as the invariant curve itself). For two caustics, D and E, we say that E is "larger" than D if D is contained in the convex hull of E. Consider the "largest" caustic for the billiard ball map on $\partial\Omega$ intersecting the (compact) invariant curve K, say E. Let l be a tangent to E at $p \in K \cap E$. As in the previous proposition, $f_\epsilon(l)$ lies outside E or tangent to E. Also $f_\epsilon(l) \in K$, and K lies inside (the convex hull of) E, so $f_\epsilon(l) \in E \cap K$. Therefore,

$$f_\epsilon^n(l) \in (E \cap K), \quad \forall n \geq 0.$$

But since the rotation number of f_ϵ on K is irrational and $f_\epsilon \in C^2(K)$, the orbit of l is dense in K and so, since f_ϵ is continuous, $E \cap K = K$, that is, K is a caustic for the billiard ball map on $\partial\Omega$.

By Proposition 13, K is also an evolute of C. □

Proof of Theorem 6. The proof proceeds by contradiction. Assume that D≠C is the caustic nearest C that is also an evolute, that is, there are no caustics in the annulus, $\mathfrak{B}$, between C and D that are evolutes of C (it is possible that D is $\partial\Omega$). By Claim 12, for some sufficiently small $\epsilon > 0$, f_ϵ has a closed invariant curve K strictly between C and D with irrational rotation number, and by Proposition 14, K is an evolute of C and also a caustic, arriving at a contradiction. □

9. Calculation of the Lazutkin Parameter

We wish to find those smooth curves which have the weaker transitivity property (and thus satisfy equation (3.1)). Let a be a simple closed strictly convex C^∞ planar curve given by its tangent angle ($0 \leq \theta \leq 2\pi$) and its curvature ($k(\theta) > 0$);

$$a(\theta) = \Big(\int_0^\theta \cos(t)\frac{dt}{k(t)}, y_a^0 + \int_0^\theta \sin(t)\frac{dt}{k(t)}\Big).$$

Let b be the Q-evolute of a ($Q > 0$), given by its tangent angle, ϕ, and curvature, v, i.e, $v = L(Q, k)$. Then for some θ_1, t_1, θ_2, and t_2

$$b(\phi) = a(\theta_1) + t_1(\cos\theta_1, \sin\theta_1) = a(\theta_2) - t_2(\cos\theta_2, \sin\theta_2),$$

and

$$Q = t_1 + t_2 - (s(\theta_2) - s(\theta_1)),$$

where s is the arclength along a, $s(\theta) = \int_0^\theta k^{-1}(t)dt$.

Since a is assumed to be an invariant curve for the billiard ball map on b,

$$\phi \ = (\theta_1 + \theta_2)/2 \ .$$

A calculation of $t_1 + t_2$ [1] shows that

$$Q = \frac{1}{\cos\Delta}\int_{-\Delta}^{\Delta} \cos(u)k^{-1}(\phi+u)du - \int_{-\Delta}^{\Delta} k^{-1}(\phi+u)du,$$

where $\Delta = \theta_2 - \phi = \phi - \theta_1$.

The right side of the equation above is clearly smooth in Δ near zero, and it turns out to vanish to third order in Δ. In fact, $Q = \Delta^3 f(\Delta)$ where f is smooth in Δ near zero and $f(0) = (12k)^{-1} > 0$.

Proposition 15. *$L(Q, k)$ is a differential operator in k which is smooth in $Q^{2/3}$ for sufficiently small Q.*

Proof (outline). The implicit function theorem shows that if $k \in C^n$, then Q is a C^n function of Δ and Δ is a C^n function of $Q^{1/3}$.

It also follows that t_1 and t_2 are C^n functions of ϕ and $Q^{1/3}$. Thus v is a C^n function of ϕ and of $Q^{1/3}$. Now observe that $L(Q, k)$ is even in Q. □

Lemma 16. *Assume that Ω is a bounded planar domain with a C^∞ and strictly convex boundary. Assume that Ω is integrable near $\partial\Omega$. Then there is a neighborhood of $\partial\Omega$ in Ω in which the caustics are C^∞ and strictly convex.*

Proof. The zero section of $S_{\partial\Omega}\Omega$ is $(s, 0)$, in terms of arc length s and the incidence angle at $\partial\Omega$.

The calculations above show that Q is continuous in the incident angle Δ, and hence the invariant circles approach $\partial\Omega$ as $Q \to 0$.

Let v denote the curvature of $\partial\Omega$. Since as $Q \to 0$ $L(Q, k) \to v$ and since there is a τ with $v(\phi) \geq \tau > 0$, it follows that $L(Q, k)$ is formally invertible. Since the caustics are assumed to exist, Proposition 15 implies that for Q sufficiently small (and $\partial\Omega$ fixed), $k \in C^\infty$ and $k > 0$. □

10. An Equation for Transitivity

Let v denote the curvature of $\partial\Omega$. Since Ω is integrable, for each R in some nonempty interval $[0, T]$, there is a caustic with curvature k s.t. $v = L(R, k)$. By Theorem 6, for each $R \in [0, T]$ and $n \geq 0$ there are $P_n = P(R, n) > 0$, $Q_n = Q(R, n)$, and $w_n = w(R, n)$ with

$$L(R, k) = v, \quad L(P_n, k) = w_n, \quad \text{and} \;\; L(Q_n, w_n) = v,$$

so that in addition $P_0 = R$, $Q_0 = 0$, and as $n \to \infty$ $P_n \to 0$ while $Q_n \to R$ ($w_n \to k$ as $n \to \infty$).

Recall from Proposition 15 that

$$L(R, k) \sim \sum_{j=0}^{\infty} L_j(k) R^{2j/3},$$

with $L_j(k)$ an ordinary differential operator. Also, $L(0, k) = k$, so L_0 is the identity.

Interpreting $L(Q_n, L(P_n, k)) = L(R, k)$ in the sense of power series, we get

$$L_0k + (L_1k)(P_n^{2/3} + Q_n^{2/3}) + (L_1^{two}k)P_n^{2/3}Q_n^{2/3} + (L_2k)(P_n^{4/3} + Q_n^{4/3}) = L_0k + L_1kR^{2/3} + L_2kR^{4/3} + O(R^{6/3}), \tag{10.1}$$

where $L_1^{two}k$ is the coefficient of $R^{2/3}$ in $L_1(k + L_1kR^{2/3})$.

As n approaches infinity,

$$R^{2/3} \sim P_n^{2/3} + Q_n^{2/3} + P_n^{2/3}g(k)(AP_n^{2/3} + BQ_n^{2/3}) + O(R^{6/3}).$$

Thus (10.1) with R approaching 0 yields $A = 0$ and

$$L_1^{two}(k(R)) = L_1(k(R))G(k(R)) + 2L_2(k(R)). \quad (G(k) = Bg(k).)$$

Since as $R \to 0$, $k(R) \to v$,

$$L_1^{two}v = L_1(v)G(v) + 2L_2v.$$

After solving for L_1 and L_2 from the geometric description in §9 this equation becomes

$$\begin{aligned} &\Big(\frac{3}{2}\Big)^{1/3}\Big(\frac{1}{60}v^{4/3}v^{(4)} - \frac{1}{30}v^{1/3}v'v^{(3)} - \frac{1}{45}v^{1/3}(v'')^2 + \frac{7}{108}v^{-2/3}(v')^2v'' \\ &\qquad + \frac{1}{20}v^{4/3}v'' - \frac{2}{81}v^{-5/3}(v')^4 - \frac{1}{36}v^{1/3}(v')^2 + \frac{1}{10}v^{7/3}\Big) \\ &\qquad = G(v)\Big(\frac{3}{2}\Big)^{2/3}\Big(\frac{1}{6}v^{2/3}v'' - \frac{2}{9}v^{-1/3}(v')^2 - \frac{1}{2}v^{5/3}\Big). \end{aligned} \tag{10.2}$$

Equation (10.2) is the desired transitivity property equation, and is satisfied by the curvature of any ellipse and its rotations. With the additional requirement that the curve be a closed convex curve (there are two conditions for closure), ellipses (a two parameter family of curves with an additional parameter for rotation) are the only curves whose curvature satisfies this equation (see [1], but it is also possible to prove this by checking the solutions' dependence on initial conditions directly by numerical means as in [2]). We record this.

Proposition 17. *The only C^∞ closed strictly convex boundaries in the plane whose curvatures satisfy* (10.2) *are ellipses.*

Proof. Using Lemma 16, Theorem 6, and Proposition 17, Theorem 1 is proved. □

NOTE. Mather [13] has shown that closed twice differentiable planar curves that are not strictly convex (a point at which the curvature vanishes exists) are not integrable. Thus the above completes the solution of Birkhoff's conjecture for C^∞ integrable planar regions (or those regions which are sufficiently differentiable so that KAM theory applies).

References

1. E. Y. Amiran, *Caustics and evolutes for convex planar domains*, J. Diff. Geom. **28** (1988), 234–234.
2. ———, *Lazutkin coordinates and invariant curves for outer billiards*, J. Math. Phys. **36** (1995), 1232–1241.
3. V. I.Arnold, *Proof of A. N. Kolmogorov's theorem on the preservation of quasi-periodic motions under small perturbations of the Hamiltonian*, Russian Math. Surveys **18** (1963), 13–40.
4. ———, *Mathematical Methods of Classical Mechanics*, Springer–Verlag, New York, NY, 1980.
5. G. D. Birkhoff, *On the periodic motions of dynamical systems*, Reprinted in Collected Mathematical Papers, II, Amer. Math. Society, Providence, RI, 1950, pp. 111–229.
6. A. Denjoy, *Sur les courbes définies par les équations différentielles á le surface de tore*, J. Math. Pure. et Appliq. **11** (1932), 333–375.
7. Yu. S. Il'yashenko and S. Yu. Yakovenko, *Nonlinear Stokes phenomena in smooth classification problems*, Advances in Soviet Mathematics **14** (1993), 235–287.
8. A. Katok, *Some remarks on Birkhoff and Mather twist map theorems*, Erg. Th. and Dyn. Sys., **2** (1982), 185–194.
9. ———, *Periodic and quasi-periodic orbits for twist maps*, In Lecture Notes in Physics, no. 179, Dynamical Systems and Chaos (L. Garrido, eds.), Springer–Verlag, New York, NY, 1983.
10. A. N. Kolmogorov, *On quasi periodic motions under small perturbations of the Hamiltonian*, Dokl. AN SSSR **98** (1954), 527–530.
11. V. F. Lazutkin, *Existence of a continuum of closed invariant curves for a convex billiard*, Math. USSR Izvestija **7** (1973), 185–214.
12. S. Marvizi and R. B. Melrose, *Spectral invariants of convex planar regions*, J. Diff. Geometry **17** (1982), 475–502.
13. J. N. Mather, *Glancing billiards*, Erg. Th. and Dyn. Sys. **2** (1982), 397–403.
14. R. Melrose, *Equivalence of glancing hypersurfaces*, Invent. Math. **37** (1976), 165–191.
15. J. Moser, *On invariant curves of area preserving mappings of an annulus*, Nachr. Akad. Wiss. Götingen Math. Phys. 1962, 1–20.
16. H. Poincaré, *Sur un théorèm de géométrie*, Randiconti del Circolo Matematico di Palermo **33** (1912).

Mathematics Department Western Washington University Bellingham, WA 98225-9063
edoh@cc.wwu.edu

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX