

Glancing billiards

JOHN N. MATHER

*Institut des Hautes Etudes Scientifiques, 35 Route de Chartres,
91440 Bures-sur-Yvette, France*

(Received 6 January 1982)

Abstract. Consider the billiard ball problem in an open, convex, bounded region of the plane whose boundary is C^2 and has at least one point of zero curvature. Then there are trajectories which come arbitrarily close to being positively tangent to the boundary and also come arbitrarily close to being negatively tangent to the boundary.

1. Statement of the result

Let R be an open convex region in the plane, and suppose ∂R is C^2 . Consider a point $p \in R$ which moves in a straight line until it hits ∂R and then bounces off according to the rule ‘angle of incidence equals angle of reflection’. The study of the resulting trajectories is often called the *billiard ball problem*.

We will say a trajectory is ε -glancing if for at least one bounce the angle of reflection (with either the positive or negative tangent of ∂R at the point of reflection) is $< \varepsilon$. If $\varepsilon < \pi/2$, we can distinguish between a *positively* ε -glancing trajectory and a *negatively* ε -glancing trajectory according to whether it is the positive or negative tangent to ∂R which the direction of reflection is close to.

A trajectory might be positively ε -glancing at one bounce and negatively ε -glancing at another bounce. Thus, one can ask whether for every $\varepsilon > 0$, there exist trajectories which are both positively and negatively ε -glancing. We will prove a theorem which shows in some cases the answer is yes.

THEOREM. *If the curvature of ∂R vanishes at some point, then for every $\varepsilon > 0$, there exist trajectories which are both positively and negatively ε -glancing.*

2. Formulation as a dynamical system

The proof depends on the formulation of the billiard ball problem in terms of area-preserving diffeomorphisms of the annulus. We begin by recalling this formulation.

Let $A = \partial R \times [-1, 1]$ and $A^0 = \partial R \times (-1, 1)$. For $x \in \partial R$ and $u \in (-1, 1)$, let $r(x, u)$ be the ray in the plane starting at x which makes the angle $\theta = \cos^{-1} u$ with the positive tangent to ∂R at x . Let $f_1(x, u)$ be the point of intersection of $r(x, u)$ with ∂R which is not x . Let $f_2(x, u) = \cos \theta_1$, where θ_1 is the angle which $r(x, u)$ makes with the positive tangent to ∂R at $f_2(x, u)$. (See figure 1.) Let

$$f(x, u) = (f_1(x, u), f_2(x, u)) \in A^0.$$

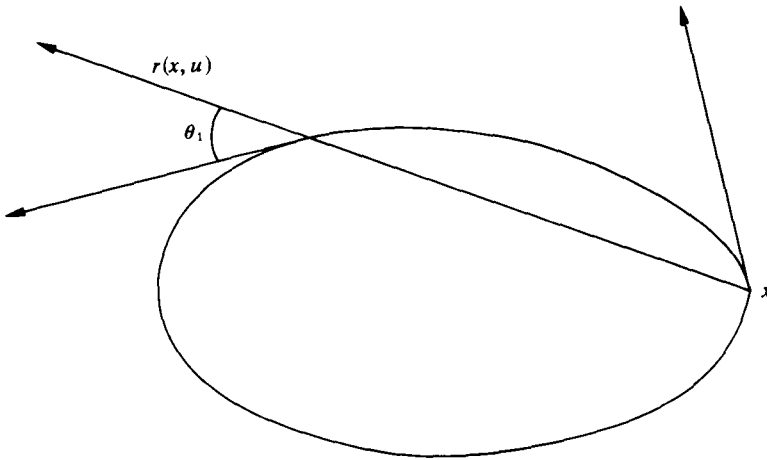


FIGURE 1

Then $f: A^0 \rightarrow A^0$ is a C^1 diffeomorphism.

The billiard ball problem is equivalent to the study of the trajectories of f .

Let ds denote the differential of the arc length on ∂R . Then f preserves the area form $ds du$. Moreover, it is easily seen that

$$\partial f_1 / \partial u < 0 \tag{1}$$

everywhere. From this, it follows that a theorem of G. D. Birkhoff [2, § 3], [1, § 44] applies to f .

3. Birkhoff's theorem

Let $f: A^0 \rightarrow A^0$ be a C^1 diffeomorphism which preserves orientation and maps each topological end of A^0 to itself. Suppose f is area preserving and (1) holds everywhere in A^0 . Let U be an open subset of A such that $\partial R \times \{-1\} \subset U$, $\bar{U} \cap (\partial R \times \{1\}) = \emptyset$, and $\partial R \times \{-1\}$ is a deformation retract of U . Suppose $fU = U$. Then the frontier of U has the form

$$\text{frontier } U = \{(x, \gamma(x)): x \in \partial R\},$$

where $\gamma: \partial R \rightarrow (-1, 1)$ is a Lipschitz mapping.

The hypothesis that f maps each topological end of A^0 to itself is equivalent to the assertion that for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$f(\partial R \times (-1, -1 + \delta)) \subset \partial R \times (-1, -1 + \epsilon)$$

and

$$f(\partial R \times (1 - \delta, 1)) \subset \partial R \times (1 - \epsilon, 1).$$

We have formulated Birkhoff's theorem according to our needs. This formulation is slightly different from either of Birkhoff's formulations ([2, § 3] and [1, § 44]). However, Birkhoff's proof still applies to give the result we have stated.

We may reduce to the case discussed in [1, § 44] by identifying $\partial R \times \{-1\}$ to a point. The image of U under this identification is a simply connected open region in the plane, and its boundary is a 'curve' as defined by Birkhoff [1, § 42]. Birkhoff

assumes differentiability hypotheses in [1, § 44] which are stronger than ours, but the proof goes through without change under our differentiability hypothesis. The conclusion that ‘the curve lies entirely within [the barred angle]’ is equivalent to our conclusion that γ is Lipschitz.

4. *Application of Birkhoff’s theorem*

From now on, we will assume that the conclusion of our theorem is false. In § 5, we will derive a contradiction. Let

$$V_- = \partial R \times [-1, -1 + \varepsilon], \quad V_+ = \partial R \times (1 - \varepsilon, 1].$$

Let

$$V = \bigcup_{n=-\infty}^{\infty} f^n(V_- \cap A^0) \cup (\partial R \times \{-1\}).$$

By our assumption that the conclusion of our theorem is false, we obtain that for $\varepsilon > 0$ sufficiently small,

$$V \cap V_+ = \emptyset.$$

From now on, we suppose ε is chosen this small. Let B be the connected component of $A \setminus V$ which contains $\partial R \times \{+1\}$. Let $U = A \setminus B$. Then $f(U) = U$, $V_- \subset U$, $U \cap V_+ = \emptyset$, and $\partial R \times \{-1\}$ is a deformation retract of U .

By Birkhoff’s theorem, there exists a Lipschitz mapping

$$\gamma : \partial R \rightarrow (-1 + \varepsilon, 1 - \varepsilon)$$

such that

$$\text{frontier } U = \{(x, \gamma(x)) : x \in \partial R\}.$$

Since $fU = U$, we have

$$f(\text{frontier } U) = \text{frontier } U,$$

so there exists a homeomorphism $g : \partial R \rightarrow \partial R$ such that

$$f(x, \gamma(x)) = (g(x), \gamma(g(x))).$$

Moreover, since f is orientation preserving and maps each side of frontier U into itself, it follows that $f|_{\text{frontier } U}$ is orientation preserving, and hence $g : \partial R \rightarrow \partial R$ is orientation preserving.

5. *Proof of the theorem*

Let x_0 be a point on ∂R where the curvative vanishes. Let $y_0 \in \partial R$. Let

$$x_n = g^n(x_0), \quad y_n = g^n(y_0).$$

From the fact that g is orientation preserving, it follows that y_0 and y_1 are monotone increasing functions of y_{-1} . In this section we will show, from the fact that y_1 is a monotone increasing function of y_{-1} and the fact that the curvative of ∂R vanishes at x_0 , that y_0 is locally a monotone decreasing function of y_{-1} , for y_{-1} in a sufficiently small neighbourhood of x_{-1} . This contradiction will prove our theorem.

For two points $v, w \in \partial R$, let $h(v, w)$ denote the Euclidean distance between them. A simple geometric argument shows that

$$\partial^2 h / \partial v \partial w > 0, \quad \text{for } v, w \in \partial R, v \neq w. \tag{2}$$

Here, the second partial derivative is defined in terms of the parameterization of ∂R by arc length. For,

$$\frac{\partial h}{\partial v}(v, w) = \cos \theta,$$

where θ is the angle between the vector $v-w$ and the positive tangent to ∂R at v (see figure 2). It is obvious that $\partial\theta/\partial w < 0$, so we obtain (2).

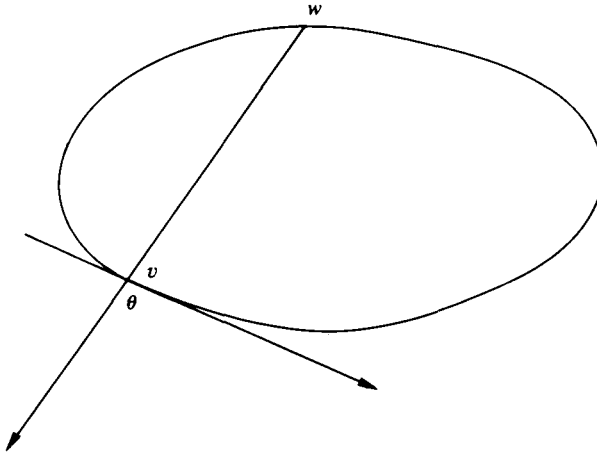


FIGURE 2. $\frac{\partial h}{\partial v}(v, w) = \cos \theta$.

From the rule ‘angle of incidence = angle of reflection’, we get

$$\frac{\partial}{\partial x_0}(h(x_{-1}, x_0) + h(x_0, x_1)) = 0. \tag{3}$$

From the fact that the curvative of ∂R vanishes at x_0 , we get

$$\frac{\partial^2}{\partial x_0^2}(h(x_{-1}, x_0) + h(x_0, x_1)) > 0. \tag{4}$$

Hence, for \bar{y} sufficiently close to x_{-1} and y' sufficiently close to x_1 , we get that there is a unique $\eta = \eta(\bar{y}, y')$ near x_0 such that

$$\frac{\partial}{\partial \eta}(h(\bar{y}, \eta) + h(\eta, y')) = 0. \tag{5}$$

From the rule ‘angle of incidence = angle of reflection’, it follows that

$$y_0 = \eta(y_{-1}, y_1), \tag{6}$$

when y_{-1} is sufficiently near to x_1 .

From (5), we get

$$\begin{aligned} \frac{\partial \eta}{\partial \bar{y}} &= -\frac{\partial^2 h}{\partial \bar{y} \partial \eta} / \left(\frac{\partial^2 h}{\partial \eta^2}(\bar{y}, \eta) + \frac{\partial^2 h}{\partial \eta^2}(\eta, y') \right) \\ \frac{\partial \eta}{\partial y'} &= -\frac{\partial^2 h}{\partial y' \partial \eta} / \left(\frac{\partial^2 h}{\partial \eta^2}(\bar{y}, \eta) + \frac{\partial^2 h}{\partial \eta^2}(\eta, y') \right). \end{aligned}$$

From (2) and (4), we then get

$$\frac{\partial \eta}{\partial \bar{y}} \Big|_{\bar{y}=x_{-1}} < 0, \quad \frac{\partial \eta}{\partial y'} \Big|_{y'=x_1} < 0. \tag{7}$$

Since y_1 is an increasing function of y_{-1} , it then follows from (6) and (7) that y_0 is a decreasing function of y_{-1} , for y_{-1} in a sufficiently small neighbourhood of x_{-1} . This contradiction proves the theorem. □

6. *A generalization*

The argument we have given concerning billiards can be generalized to certain C^1 diffeomorphisms $f: A \rightarrow A$ of the annulus $A = S^1 \times [0, 1]$. These are the area-preserving diffeomorphisms which preserve orientation, map each boundary component into itself and satisfy the following condition.

DEFINITION. Let $f: A \rightarrow A$ be a C^1 diffeomorphism of the annulus which is orientation preserving and maps each boundary component into itself. We will say f satisfies a monotone twist condition if (1) holds, where s denotes the angular coordinate on S^1 , u denotes the standard coordinate on the unit interval $[0, 1]$, and $f_1(s, u)$ denotes the s -coordinate of $f(s, u)$.

When f satisfies these conditions and preserves the area form $ds du$, it is possible to introduce further conditions which guarantee the non-existence of invariant circles. These conditions are based on the observation that f can be defined by a generating function. To be precise, let $A = \mathbb{R} \times [0, 1]$ denote the universal covering space of A and let $\tilde{f}: \tilde{A} \rightarrow \tilde{A}$ be a lift of f . Let $\tilde{f}_i: \mathbb{R} \rightarrow \mathbb{R}$ be given by

$$\tilde{f}_i = \tilde{f}|_{\mathbb{R} \times i}, \quad i = 0, 1.$$

Let

$$B = \{(x, y) \in \mathbb{R}^2: \tilde{f}_0(x) \leq y \leq \tilde{f}_1(x)\}.$$

Then there is a C^2 function h on B such that

$$(s_1, u_1) = f(s, u)$$

is uniquely defined by

$$u = \frac{\partial h}{\partial s}(s, s_1) \quad u_1 = -\frac{\partial h}{\partial s_1}(s, s_1). \tag{8}$$

Of course, such generating functions are well known in classical mechanics, and this form of the generating function was explicitly defined, e.g., in [5]. The fact that $(s_1, u_1) = f(s, u)$ can be defined by (8) is equivalent to the hypotheses we imposed above on f .

Birkhoff's theorem still applies to this more general situation. So, suppose $\gamma: S^1 \rightarrow (0, 1)$ is a Lipschitz function whose graph is invariant under f . Let $g: S^1 \rightarrow S^1$ be the homeomorphism such that

$$f(x, \gamma(x)) = (g(x), \gamma(g(x))).$$

In view of the fact that f is C^1 and γ is Lipschitz, it follows that g is Lipschitz. Let

$$s \in S^1, \quad s_1 = g(s), \quad s_{-1} = g^{-1}(s).$$

From (8), we get

$$\frac{\partial}{\partial s}(h(s_{-1}, s) + h(s, s_1)) = 0.$$

Differentiating formally with respect to s , we get

$$h_{12}(s_{-1}, s) \frac{ds_{-1}}{ds} + h_{12}(s, s_1) \frac{ds_1}{ds} = -(h_{22}(s_{-1}, s) + h_{11}(s, s_1)). \quad (9)$$

Since s_{-1} and s_1 may only be Lipschitz functions of s , the derivatives ds_{-1}/ds and ds_1/ds may not exist everywhere, but they exist almost everywhere, and Birkhoff's theorem gives positive lower and upper bounds for them. Thus, (9) is a necessary condition on the circle. Since $h_{12} > 0$ everywhere, and $(ds_{-1}/ds), (ds_1/ds) > 0$ everywhere, we must have

$$h_{22}(s_{-1}, s) + h_{11}(s, s_1) < 0,$$

for all $s \in S^1$, if the circle exists.

In the case of the billiard ball problem, we showed that when s is the point on the boundary where the curvature is 0, then

$$h_{22}(\bar{s}, s) + h_{11}(s, s') > 0$$

for any \bar{s}, s' , and so obtained a contradiction.

7. Related literature

According to Lazutkin [4], if ∂R is C^{55} and its curvature never vanishes, then there exist caustics for the billiard ball problem, and consequently there cannot exist trajectories which are both positively and negatively ε -glancing. A general discussion of the billiard ball problem in a convex region is contained in [7, pp. 86–89]. Bunimovich [3] proved that billiards in a stadium satisfy the Bernoulli property. Of course, our hypothesis does not imply his hypothesis, so our result does not follow from his. But for the systems he considers, his result is much stronger than ours. His result raised the question which led to our result.

More generally, we would like to find necessary and sufficient conditions for the existence of invariant circles, as discussed in § 6. There, we discussed conditions which are easily proved to be necessary. Newman and Percival [6] have discovered numerically what appear to be necessary conditions. These are probably related to ours, but much more work is necessary to make the relation clear.

REFERENCES

- [1] G. D. Birkhoff. Surface transformations and their dynamical applications. *Acta Math.* **43** (1922), 1–119. Reprinted in *Collected Mathematical Papers*, vol. II. Amer. Math. Soc.: New York, 1950, pp. 111–229.
- [2] G. D. Birkhoff. Sur quelques courbes fermées remarquables. *Bull. Soc. Math. de France* **60** (1932), 1–26. Reprinted in *Collected Mathematical Papers*, vol. II. Amer. Math. Soc.: New York, 1950, pp. 418–443.
- [3] L. A. Bunimovich. On the ergodic properties of nowhere dispersing billiards. *Commun. Math. Phys.* **65** (1979), 295–312.

- [4] V. F. Lazutkin. The existence of caustics for a billiard problem in a convex domain. *Math. USSR Izvestija* 7 (1973), 185–214. Translation of *Izvestija, Mathematical Series*, Academy of Sciences of the USSR, 37, 1973.
- [5] J. N. Mather. *Existence of quasi-periodic orbits for twist homeomorphisms of the annulus*. (To appear in *Topology*.)
- [6] R. P. A. C. Newman & I. C. Percival. Bounds on domains in phase space for Lagrangian systems with discrete time. Preprint, Queen Mary College.
- [7] Ya. G. Sinai. *Introduction to Ergodic Theory*, Math. Notes, Princeton University Press: Princeton 1977. (Translated from Russian by V. Scheffer.)