

## THE BENETTIN–STRELCYN OVAL BILLIARD REVISITED

M. HÉNON

*C.N.R.S., Observatoire de Nice, France*

and

J. WISDOM

*Observatoire de Nice, France\**

Received 15 December 1982

The wall of the oval billiard defined by Benettin and Strelcyn has a discontinuous curvature; the associated mapping has a discontinuous derivative. We find that invariant curves of the mapping are generally destroyed when they cross a line of discontinuity. They can survive, however, in particular cases where two successive discontinuities cancel each other.

### 1. Introduction

Benettin and Strelcyn [1] investigated numerically the motion of a point mass in an oval billiard (whose exact definition will be recalled below), using an appropriate two-dimensional surface of section. As a parameter  $\delta$  is varied from 0 to 1, a continuous transition from an ergodic system to an integrable system is observed. At first there exists only one large chaotic region; it then divides into a growing number of distinct chaotic regions, with different Lyapunov numbers.

Benettin and Strelcyn focused their attention exclusively on the chaotic regions. The present paper is in a sense complementary: we shall examine the boundaries between the chaotic regions, which presumably are invariant curves in the surface of section. Another motivation for this study is as follows. Most mappings and dynamical systems studied in the literature are described by

equations which are differentiable a sufficient number of times for the KAM theorem to be applicable; as is well known, this theorem accounts quite well for many observed features. We shall call this the *regular case*. In the oval billiard, however, the wall has discontinuities at four points in its curvature (defined as the inverse of the radius of the osculating circle). This induces discontinuities in the first-order derivatives of the mapping. As a result, the KAM theorem does not apply. From a theoretical point of view, one would like to know what happens in such a case, and how the usual properties are altered. The results might also be of practical interest since discontinuities are bound to exist in many real situations. Mappings with discontinuities in various derivatives have been previously studied by Froeschle [5, 6] and Brahic [3].

Benettin and Strelcyn found that, as far as the chaotic regions are concerned, the oval billiard exhibits no obvious difference with the regular case. It will be seen below that this is not so for the invariant curves; their properties are markedly different from those of the regular case.

\* Present address: Department of Physics, University of California, Santa Barbara, California 93106, USA.

### 2. Definition and known properties

The oval billiard is defined as follows. In a rectangular system of coordinates  $(x, y)$ , we consider the four points  $P_1, P_2, P_3, P_4$  with coordinates  $x = \pm 1, y = \pm 1$ , forming a square, and we join them by four circular arcs in such a way that the arcs have a common tangent at their meeting points (fig. 1). The centres  $O_1$  to  $O_4$  of the arcs lie on the  $x$  and  $y$  axes; the billiard is symmetric with respect to these axes and is characterized by a single parameter  $\delta$ , the distance from  $O_1$  to the side  $P_1P_2$  of the square. A point mass moves in this oval billiard, bouncing specularly off the walls.

This is a Hamiltonian system with two degrees of freedom. A natural and convenient surface of section is introduced by restricting attention to the successive collisions of the particle with the billiard wall. In order to define coordinates on this surface of section, we consider the curvilinear distance  $\psi$  along the wall from some origin  $B$  to the collision point  $A$ , and the oriented angle  $\alpha$  from the normal to the incoming trajectory (fig. 2). For  $B$  we take the intersection of the billiard wall with the positive  $x$  axis. Following [1], we shall use as actual coordinates  $\eta = \psi / \Psi$ , where  $\Psi$  is the total length of the wall, and  $s = \sin \alpha$ .  $\eta$  is a cyclic variable, defined modulo 1, while  $s$  is restricted to the interval  $[-1, +1]$ ; therefore the surface of section is essentially a segment of a cylinder. By cutting this

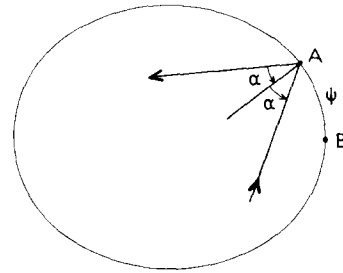


Fig. 2. Definition of the coordinates  $\Psi$  and  $\alpha$ .

cylinder along the line  $\eta = 0$ , one can represent it as a rectangle in the  $(\eta, s)$  plane, with

$$0 \leq \eta < 1, \quad -1 \leq s \leq 1. \tag{1}$$

The motion of the particle in the billiard corresponds to a mapping  $T$  of the surface of section onto itself.  $T$  can be shown to be area-preserving [2].

The oval billiard has three symmetries, corresponding respectively to a change in sign of  $x, y$ , or the time  $t$ . In the surface of section, these symmetries correspond to

$$x' = -x : \eta' = 1/2 - \eta \pmod{1}, \quad s' = -s, \tag{2}$$

$$y' = -y : \eta' = 1 - \eta \pmod{1}, \quad s' = -s, \tag{3}$$

$$t' = -t : \eta' = \eta, \quad s' = -s. \tag{4}$$

For  $\delta = 1$ , the oval billiard reduces to a circle, and the system is integrable:  $\alpha$  is constant. The  $(\eta, s)$  surface of section is filled with invariant curves, which are the horizontal lines  $s = \text{constant}$ . For  $\delta = 0$ , the two arcs  $P_1P_2$  and  $P_3P_4$  become semicircles and the two arcs  $P_2P_3$  and  $P_4P_1$  become straight line segments. This particular case is known as a *stadium* (stadions are more generally defined with  $P_1P_2P_3P_4$  an arbitrary rectangle); and stadions have been shown to be ergodic [4]. The points of a trajectory generally fill densely the surface of section. In varying  $\delta$  from 0 to 1, Benettin and Strelcyn [1] observed a gradual transition from the extreme case of an ergodic system, corresponding to complete disorder, to the other

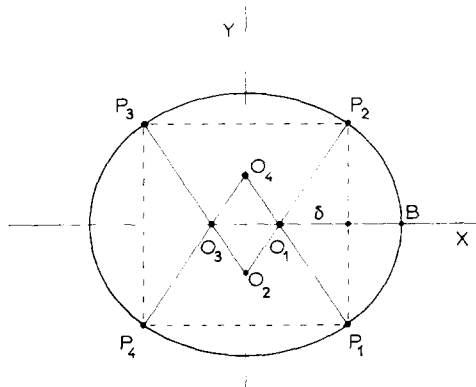


Fig. 1. Definition of the oval billiard.

extreme case of an integrable system, corresponding to complete order. The intermediate cases,  $0 < \delta < 1$ , appear from numerical results to belong to the general “mixed” case: the surface of section is filled with a mixture of invariant curves and chaotic trajectories.

The area occupied by invariant curves grows regularly as  $\delta$  increases from 0 to 1. At first, the invariant curves form a number of “islands”, lying in a chaotic “sea” which forms a single connected region. Fig. 3 represents for  $\delta = 0.6$  a single orbit with  $N = 10^7$ ;  $N$  is the number of iterations of the mapping  $T$ , or the number of collisions. (Plotting several million points for a chaotic orbit in their order of appearance would take an impossibly long time on a mechanical plotter. A special routine was devised for the figures of this paper: the surface of section is divided into  $512 \times 512$  little square cells; the dimension of a cell is approximately equal to the diameter of the pen. Each cell is mapped by a bit in computer memory. Occupied cells are noted as the computation progresses. Plotting starts only after the computation has ended: the cells are read from top to bottom and back, column by column. Two successively filled cells are connected by a line; the pen is lifted only to avoid an unfilled cell. A

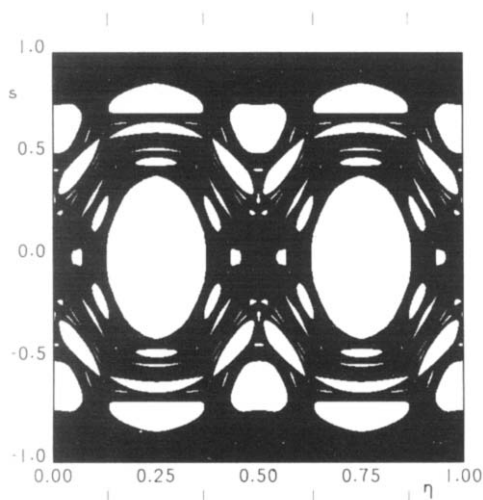


Fig. 3. The chaotic region for  $\delta = 0.6$  (cf. fig. 7 in [1]). Number of iterations:  $N = 10^7$ .

typical plotting time with this method is 15 minutes).

The orbit fills the chaotic region; most cells are occupied, producing an almost uniform black region. The white holes correspond to islands, which appear here as the complement of the chaotic region. Fig. 4 shows a number of orbits started inside the islands; the usual structure of concentric invariant curves is observed. (Actually there are also chaotic regions inside the islands, in the form of thin annular bands; we do not consider these here).

The three symmetries (2) to (4) of the billiard are manifest on figs. 3 and 4. Because of these symmetries, the region of the surface of section defined by

$$0 \leq \eta \leq \frac{1}{4}, \quad 0 \leq s \leq 1 \quad (5)$$

is sufficient to characterize the whole. From now on we shall represent only this region, in order to show more detail.

As  $\delta$  is increased, Benettin and Strelcyn discovered a striking phenomenon. At a value of  $\delta$  between 0.75 and 0.76, the chaotic “sea” suddenly appears to divide into two parts. Figs. 5 and 6 show these two components for  $\delta = 0.76$ . Each figure represents a single orbit, with 3 and 10

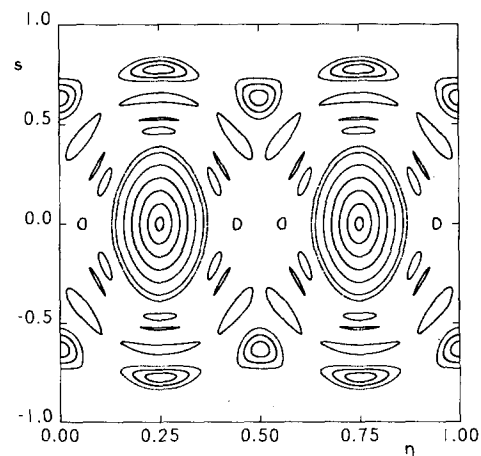


Fig. 4. Islands for  $\delta = 0.6$  (cf. fig. 7 in [1]).

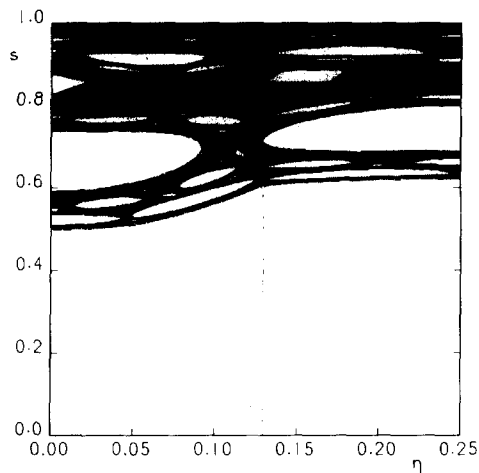


Fig. 5. The upper chaotic region for  $\delta = 0.76$  (cf. fig. 10 in [1]).  $N = 3 \times 10^6$ .

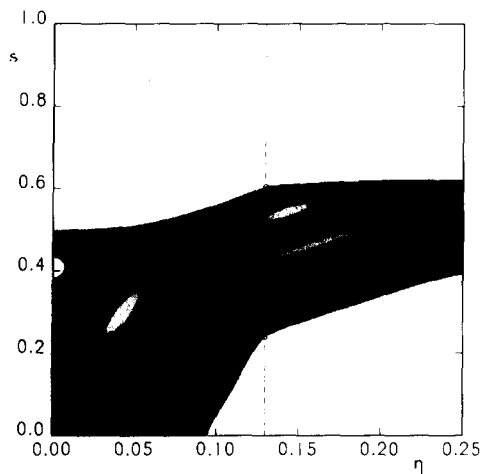


Fig. 6. The lower chaotic region for  $\delta = 0.76$  (cf. fig. 10 in [1]).  $N = 10^7$ . The white region occupying the lower right is part of a large island surrounding the periodic orbit at  $\eta = 0.25$ ,  $s = 0$ .

million computed points respectively. So we have here apparently two chaotic regions, which do not communicate. They are separated by a sharply defined boundary; by superimposing the two figures, one finds that the lower boundary in fig. 5 coincides with the upper boundary in fig. 6 to the resolution of the pictures.

These observations suggest that there exists an invariant curve between the two chaotic regions.

This curve differs from those found so far in the islands (fig. 4) in that it runs continuously from the left side of the full surface of section ( $\eta = 0$ ) to the right side ( $\eta = 1$ ); we shall call it a *transverse invariant curve*. For larger values of  $\delta$ , more such curves seem to come into existence. Figs. 7 and 8 correspond to the case  $\delta = 0.85$ . Each figure represents now three different orbits; 2 million points have been computed for each orbit. Thus, going from top to bottom, one counts now six distinct

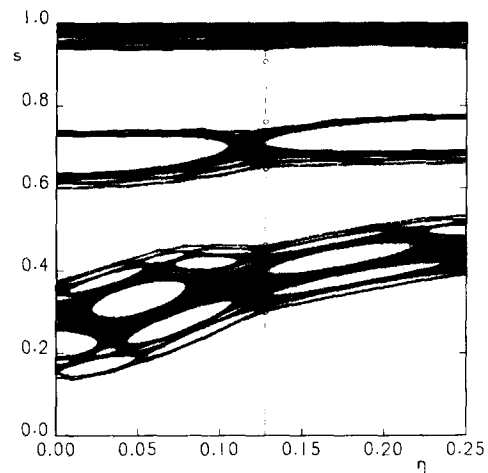


Fig. 7. Three chaotic regions for  $\delta = 0.85$  (cf. fig. 11 in [1]).  $N = 2 \times 10^6$  for each orbit.

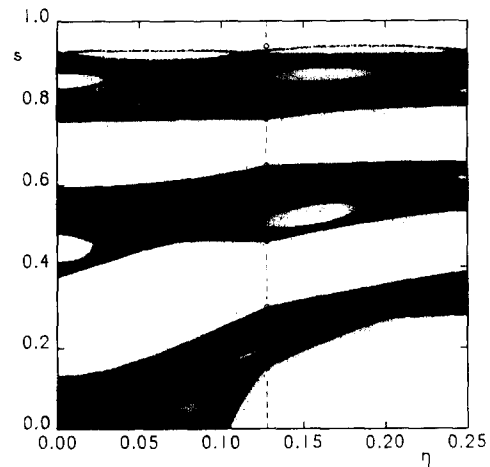


Fig. 8. Three other chaotic regions for  $\delta = 0.85$  (cf. fig. 11 in [1]).  $N = 2 \times 10^6$  for each orbit.

chaotic regions, which are displayed alternatively on figs. 7 and 8 for clarity. This suggests the existence of at least five transverse invariant curves.

As  $\delta$  continues to increase and approaches 1, the chaotic regions become thinner and more numerous; they also tend to the shape of horizontal bands. This suggests that the number of transverse invariant curves also increases monotonically, until in the limit  $\delta = 1$  they fill the whole surface of section.

### 3. Discontinuities

This mixture of invariant curves and chaotic regions, with the invariant curves becoming ever more abundant as an integrable case is approached, is quite similar to the familiar picture of the regular case, observed in many examples and described by the KAM theorem. However, one important difference can be noticed. According to the KAM theorem, when one perturbs an integrable system, many invariant curves survive, with their shape simply distorted; these surviving curves form a set of positive measure. In the oval billiard, therefore, we would expect for  $\delta < 1$  a set of transverse invariant curves of positive measure. On the contrary, numerical results suggest that only a finite number of isolated transverse invariant curves survive. (We are not considering here what happens inside the islands.)

This incomplete agreement is not surprising, because one of the conditions of the KAM theorem is that the mapping should be derivable a sufficient number of times. This condition is not satisfied by the oval billiard. The curvature of the wall is discontinuous at the four points  $P_i$ . In general, discontinuities in the  $n$ th derivative of the billiard wall result in discontinuities in the  $(n - 1)$ th derivative of the mapping itself. Therefore our mapping has discontinuities in its first derivative; it belongs to class  $C^0$  but not to class  $C^1$ .

We call  $\eta_i$  the normalized coordinate of point  $P_i$ . If a point  $Q$  moves in the  $(\eta, s)$  surface of section,

the mapping of  $Q$  has a discontinuous first derivative whenever one of the four vertical lines  $\eta = \eta_i$  is crossed. Inside each of the four regions delimited by these lines (which we shall call *lines of discontinuity*), however, the mapping is analytic. For  $\delta = 0.6$ , for instance, there is  $\eta_2 = 0.133394\dots$ ; the positions of the four lines of discontinuity are indicated by the tick marks above and below the frame in fig. 3.

A remarkable phenomenon can then be noticed: the islands seem to refrain from crossing the lines of discontinuity, although they often touch them. Similar observations can be made in figs. 5 to 8, where the line of discontinuity  $\eta = \eta_2$  is shown as a dashed line. (The only flagrant exception is the large island centered in  $\eta = 0.25$ ,  $s = 0$ , at the bottom right of figs. 6 and 8.) This has two interesting consequences. First, each island stays within a regular region, where it never meets any discontinuity. As a result, the islands should exhibit the usual picture of the regular case. This is indeed what is found: most initial conditions inside islands apparently result in invariant curves (fig. 4).

Second, the fact that in many cases the observed outer boundary of an island touches a line of discontinuity suggests that it is difficult for an invariant curve to survive the crossing of such a line. An explanation is suggested by the following heuristic argument. Suppose that an invariant curve intersects a line of discontinuity at a point  $Q_0$  and that it is differentiable at  $Q_0$ . Let  $Q_1, Q_2, \dots$  be the successive images of  $Q_0$ . Since the mapping has a discontinuous first derivative at  $Q_0$ , the invariant curve will have an angle at  $Q_1$ . This is then carried over to all points  $Q_i$ , which will also correspond to angles in the curve. But an invariant curve is characterized by an irrational rotation number, and the set of the points  $Q_i$  covers the curve densely. Therefore the invariant curve must be nowhere differentiable. (A similar argument applies more generally to mappings with discontinuities in the  $n$ th derivative). Apparently this nowhere-differentiable pathology is more than can be endured by invariant curves; and therefore islands are limited by the lines of discontinuity.

Examples of nowhere-differentiable invariant curves have been found in a mapping by Froeschlé [5]; but that mapping is a special case (see Appendix).

#### 4. Cancellation orbits

How then do we explain that some transverse invariant curves apparently manage to exist, in spite of the fact that they cross all four lines of discontinuity? And how do we explain that these curves, as defined by the boundaries of the chaotic regions, seem to be rather smooth (except where they cross the lines of discontinuity, and perhaps at a few other points), instead of being nowhere differentiable? (See figs. 5 to 8).

This second observation gives us the key. If an angle appears at the line of discontinuity but is not seen afterwards, then the discontinuity must have been cancelled in some way. This is indeed possible if the orbit happens to visit successively two points of discontinuity  $P_i$  and  $P_j$ : the second discontinuity may then cancel the effect of the first. Orbits with this property will be called *cancellation orbits*.

The simplest case of a cancellation orbit is when  $P_i$  and  $P_j$  are two successive collisions. More generally,  $P_i$  and  $P_j$  can be separated by  $n - 1$  intermediate collisions; we shall call this a cancellation orbit of order  $n$ . The discontinuity created at  $P_i$  is then cancelled only when  $P_j$  is reached, and the intermediate collisions will correspond to angles in the invariant curve. It can be shown that there are at least  $4n - 1$  cancellation orbits of order  $n$ ; the set of cancellation orbits is therefore infinite. Some of the simplest cancellation orbits, arbitrarily named a to g, are represented on fig. 9; a and b are of order 1 while c to g are of order 2. In case g, the orbit comes back over itself after a perpendicular collision, and both  $P_i$  and  $P_j$  are in  $P_2$ .

The collision of the cancellation orbit a at point  $P_2$ , for  $\delta = 0.76$ , is represented on figs. 5 and 6 by an open circle (partially hidden by the chaotic region), with coordinates  $\eta = \eta_2 = 0.129633 \dots$  and  $s = \sin \alpha = 0.605083 \dots$ . It can be seen that this

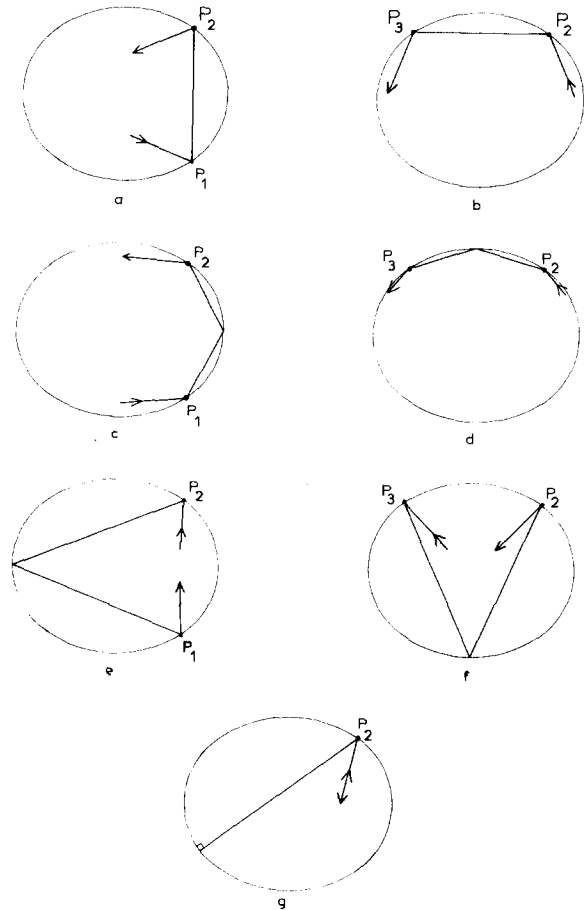


Fig. 9. Some cancellation orbits.

point lies exactly on the observed boundary between the two chaotic regions. This has been verified to an accuracy of  $10^{-5}$ .

Similarly, the cancellation orbit g is represented on fig. 6 by an open circle, with coordinates  $\eta = 0.129633 \dots$ ,  $s = 0.239900 \dots$ . It corresponds exactly to the lower boundary of the chaotic region. The invariant curve whose existence is suggested by this boundary is not a transverse invariant curve; but it surrounds the main island (centered at  $\eta = 0.25$ ,  $s = 0$ ) and crosses two lines of discontinuity, so that similar properties are observed.

The positions of the six cancellation orbits a to f for  $\delta = 0.85$  have been indicated on fig. 7;

their common abscissa is  $\eta_2 = 0.127764\dots$  and their ordinates are, in increasing order:  $s_f = 0.300150\dots$ ,  $s_e = 0.462264\dots$ ,  $s_a = 0.647648\dots$ ,  $s_b = 0.761939\dots$ ,  $s_c = 0.907647\dots$ ,  $s_d = 0.938599\dots$ . A close agreement with the apparent boundary between two chaotic regions can be observed for orbits f, e, a, d. The same six orbits, plus orbit g (with ordinate  $s_g = 0.149991\dots$ ), are represented on fig. 8; agreement is observed for orbits g, e, f, a, b. This can be taken as a confirmation of the role played by cancellation orbits. Note that there was no reason to expect agreement for all points, since the selection of the cancellation orbits a to g was somewhat arbitrary.

Up to now our concept of “discontinuity cancellation” has been rather vague; we shall try now to

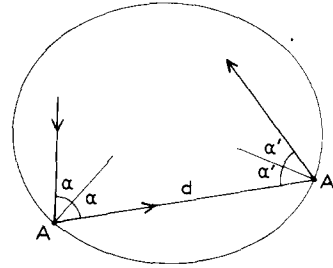


Fig. 10. Two successive collisions.

make it more definite. We consider two successive points of collision A and A' (fig. 10) and the corresponding points Q and Q' in the  $(\eta, s)$  surface of section. The mapping of the vicinity of Q into the vicinity of Q' is described to first order by the Jacobian of  $T$ , which is easily found to be

$$J = \frac{\partial(\eta', s')}{\partial(\eta, s)} = \begin{pmatrix} \frac{Cd - \cos \alpha}{\cos \alpha'} & -\frac{d}{\Psi \cos \alpha \cos \alpha'} \\ \Psi(C \cos \alpha' + C' \cos \alpha - CC'd) & \frac{C'd - \cos \alpha'}{\cos \alpha} \end{pmatrix}. \quad (6)$$

Here  $d$  is the distance  $AA'$ , and  $C$  and  $C'$  are the curvatures of the wall in A and A'.

We let now Q trace a smooth curve  $\Gamma$  in the surface of section. This generates a one-parameter family of orbits. Q' traces a curve  $\Gamma'$ . According to (6), the slopes  $p = ds/d\eta$  and  $p' = ds'/d\eta'$  of  $\Gamma$  and  $\Gamma'$  are related by

$$p' = \frac{(C'd - \cos \alpha')p \cos \alpha' + \Psi(C \cos \alpha' + C' \cos \alpha - CC'd) \cos \alpha \cos \alpha'}{-dp/\Psi + (Cd - \cos \alpha) \cos \alpha}. \quad (7)$$

As long as A and A' are ordinary points, (6) and (7) are well defined. However, if either A or A' (or both) passes through one of the points  $P_i$ , the corresponding curvature of  $C$  or  $C'$  changes discontinuously. The slope  $p'$  given by (7) is therefore discontinuous: an angle is produced in the curve  $\Gamma'$  at that point.

We label now the successive points of an orbit  $Q_0, Q_1, Q_2, \dots$  and we assume that the family of orbits includes a cancellation orbit of type a (we choose this type for definiteness; a similar argument could be applied to the other types). More precisely, for one particular orbit of the family, the two collision points  $A_1$  and  $A_2$  coincide with the junction points  $P_1$  and  $P_2$ . This will produce discontinuities in three successive mappings: from  $Q_0$  to  $Q_1$ , from  $Q_1$  to  $Q_2$ , and from  $Q_2$  to  $Q_3$ . We consider the product of the three mappings:  $Q_3 = T^3(Q_0)$ . The corresponding Jacobian is  $J_2 J_1 J_0$ ; it can be computed from (6). From it we obtain the slope  $p_3$  as a function of the slope  $p_0$ . The values of  $p_3$  on both sides of the cancellation orbit will generally be different:  $\Gamma_3$  has an angle. We ask now if it is possible for this angle to vanish. A straightforward computation shows that this happens for exactly one value of the slope  $p_0$ . When the smooth curve  $\Gamma_0$  has this particular slope (at the cancellation orbit),  $\Gamma_1$  and  $\Gamma_2$  have angles, but  $\Gamma_3$  is again smooth. All further curves  $\Gamma_i$ , with  $i > 3$ , are also smooth. We shall not quote here the value of  $p_0$ , but rather the corresponding

values of  $p_2$  which can be more conveniently compared with observations. They are, for  $\eta < \eta_2$  and  $\eta > \eta_2$  respectively:

$$(p_2)_{\text{left}} = \frac{1}{2} \Psi(1 - 2\delta + [4(1 - \delta)^2 + 1]^{1/2}) / (1 + \delta^2), \quad (8)$$

$$(p_2)_{\text{right}} = \frac{1}{2} \Psi(-1 + [4(1 - \delta)^2 + 1]^{1/2}) / (1 + \delta^2). \quad (9)$$

For  $\delta = 0.76$ , these slopes are  $1.666352 \dots$  and  $0.308913 \dots$ , respectively. Straight half-lines with these slopes are represented on fig. 11 as half-lines emanating from the point  $Q_2$  of the cancellation orbit; the figure is a magnification of the surface of section in the vicinity of  $Q_2$ . The cancellation orbit itself is represented by large dots;  $10^5$  points have been computed, starting from  $Q_2$ . A very good agreement is apparent. This indicates that the points of the cancellation orbit arrange themselves on an invariant curve which has precisely the slope required for discontinuity cancellation. As a result, an invariant curve presents only four angles, at the four abscissas  $\eta = \eta_1, \dots, \eta_4$ , and is smooth elsewhere. Fig. 11 thus provides new evidence in favor of our conjecture that the dividing invariant curves correspond to cancellation orbits.

One intriguing question is raised, though: how does the cancellation orbit manage this feat? In other words, why does it have precisely the slope which is required for its very existence?

We remark in passing that the above computation does not depend on the symmetries of the oval billiard; it could be applied to any billiard with curvature discontinuities. In particular, it is not necessary for the two curvature jumps to be equal.

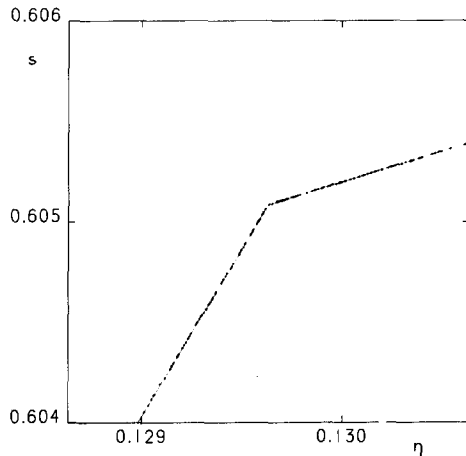


Fig. 11. Dots: cancellation orbit for  $\delta = 0.76$ . Straight lines: predicted slopes for discontinuity cancellation, given by (8) and (9).  $N = 10^5$ .

## 5. Transverse invariant curves

Thus the numerical evidence suggests that transverse invariant curves, when they exist, correspond to cancellation orbits. The converse is not true: a cancellation orbit does not necessarily form an invariant curve. For a given value of  $\delta$ , there are

an infinite number of cancellation orbits, but numerical results indicate that only a finite number of them produce an invariant curve.

These curves therefore form a subset of measure zero in the surface of section. They cannot be detected in a numerical exploration where initial conditions are randomly chosen; this explains why in [1] only chaotic regions and islands were observed. This situation is at variance with the regular case, where by picking arbitrary initial conditions one usually finds many invariant curves. Here, in order to find the transverse invariant curves, we must look specifically for them. Fortunately no trial-and-error search is necessary, since we know that they correspond to cancellation orbits: we simply take an initial point with  $\eta = \eta_2$  and the appropriate value of  $s$ . For the cancellation orbit of type a, for instance,  $s$  is given by

$$s_a = \delta(1 + \delta^2)^{-1/2}. \quad (10)$$

Similar explicit formulas for  $s$  are easily found for the types b to g of fig. 9.

For the remainder of this paper, we shall consider only the cancellation orbit of type a; similar

results should hold for other types. A typical example of an invariant curve corresponding to this cancellation orbit is shown in fig. 12 for  $\delta = 0.83$ . 5000 iterations of the mapping have been computed; the points are so dense that they give the illusion of a continuous curve. A small part of the curve is shown in fig. 13 under high magnification; here  $5 \times 10^5$  iterations have been computed. Even at that magnification, corresponding to a vertical resolution of  $2 \times 10^{-10}$ , the points appear to lie exactly on a curve.

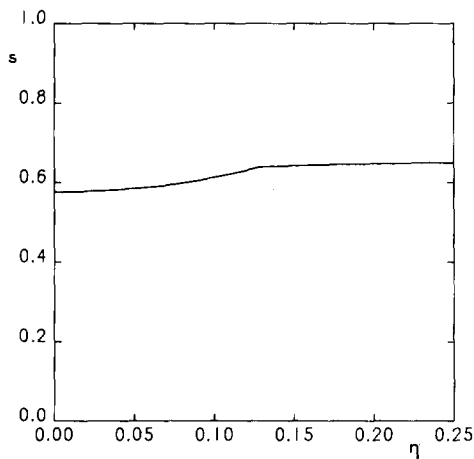


Fig. 12. The cancellation orbit of type a, for  $\delta = 0.83$ .  $N = 5000$ .

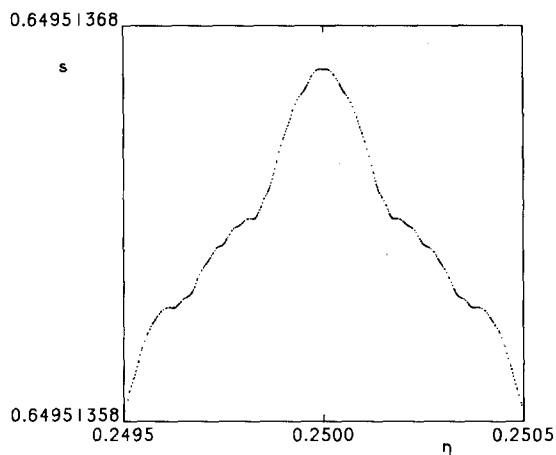


Fig. 13. Enlargement of a small part of fig. 12.  $N = 5 \times 10^5$ .

This curve appears to be differentiable (no angles are apparent), but it is very bumpy. A tentative explanation of this fact is that the discontinuity cancellation operates only at first order, not for second and higher orders. A crossing of a line of discontinuity produces then a discontinuity in the second derivative, which is propagated everywhere along the curve in the manner described in section 3. It follows that the curve is nowhere twice differentiable; in other words, it nowhere has a definite curvature. Examples of curves with this property can be seen in [6].

However, the cancellation orbit a does not always give rise to an invariant curve, even for  $\delta$  larger than 0.76, the critical value found by Benettin and Strelcyn. In many cases, the points do not form a line, but wander away from their initial location more or less rapidly and invade one of the neighbouring chaotic regions. This frequently happens in three successive phases. The points form first a chain of “blobs”; this is illustrated in fig. 14 for  $\delta = 0.839$ . This structure resembles somewhat the usual chains of islands; however, each blob has a chaotic structure, as shown by the enlargement of one of them in fig. 15. In a second phase, the blobs become joined to form a thin chaotic band. Finally, the orbit expands into a large chaotic region.

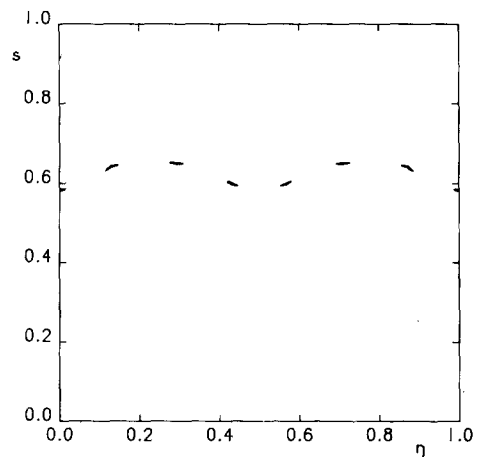


Fig. 14. The cancellation orbit of type a, for  $\delta = 0.839$ .  $N = 2500$ .

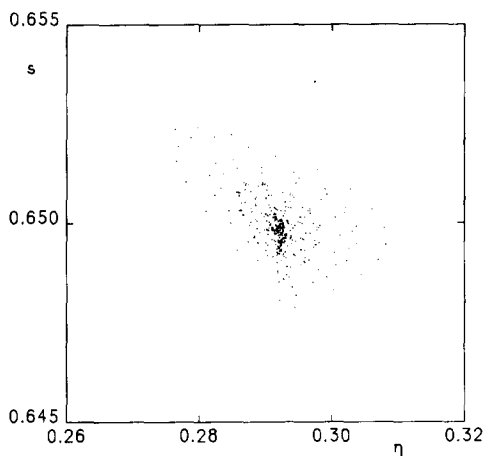


Fig. 15. Enlargement of part of fig. 14.

The time scales associated with these three phases can vary widely for different values of  $\delta$ .

Thus, the “cancellation orbit” property appears to be a necessary condition, but not a sufficient one, for the existence of a transverse invariant curve. This is easily understood. Even in the case of a sufficiently derivable mapping, for which the KAM theorem applies, an invariant curve will exist at a given location only under two conditions: (i) the perturbation relative to an integrable case should not be too large; (ii) the rotation number should not be too close to a rational. These conditions should also apply in the present problem; an invariant curve must satisfy them, in addition to being a cancellation orbit. The first condition clearly corresponds to the fact that transverse invariant curves are observed only above a critical value of  $\delta$ , i.e. sufficiently close to the integrable case  $\delta = 1$ . The second condition means that even above this critical value, a transverse invariant curve will not form if the rotation number is too close to a rational. For example, the non-existence of the invariant curve for  $\delta = 0.839$  (figs. 14 and 15) is probably a consequence of the fact that for the nearby value  $\delta = 0.8370175289\dots$ , the cancellation orbit is periodic with period 7, rotation number  $\frac{2}{7}$ . Invariant curves are therefore destroyed in a finite interval around this value of  $\delta$ .

As a matter of fact, even in the case  $\delta = 0.76$ , for which figs. 5 and 6 seemed to provide strong evidence of the existence of a transverse invariant curve, close examination reveals that this curve does not really exist. An evolution similar to that depicted above for  $\delta = 0.839$  (figs. 14 and 15), but less obvious, takes place for the cancellation orbit. The points form first a chain of 343 tiny blobs, which later join to form a thin chaotic band with a vertical thickness of the order of  $10^{-5}$ . This thickness is too small to be visible on fig. 11, thus giving the illusion that the points lie on a curve. After about 220 000 iterations, however, the third phase is reached and the points invade the lower chaotic region represented in fig. 6.

One can try to find invariant curves by choosing values of  $\delta$  for which the rotation number is as far as possible from the rationals. In the range  $0.7 < \delta < 1$ , the rotation number varies roughly from  $\frac{1}{3}$  to  $\frac{1}{4}$ . Therefore a good value for the rotation number should be the continued fraction

$$\frac{1}{3 + \frac{1}{1 + \frac{1}{1 + \dots}}} = \frac{5 - \sqrt{15}}{10}, \quad (11)$$

closely related to the golden mean [7]. The value of  $\delta$  for which the rotation number has the value (11) is found by trial and error to be  $\delta = 0.87637677\dots$ . Fig. 16 shows a highly magnified section of the corresponding cancellation orbit; this can be compared with fig. 13. Again the points appear to fall precisely on a curve, as expected. There is a noticeable difference, however: the curve of fig. 16 is very regular, and exhibits none of the bumpiness of fig. 13. It may be that the selection of this particular value of  $\delta$  ensures not only the existence of the curve but its derivability to an order higher than the first. We have no explanation to offer for this phenomenon.

In conclusion, it appears that a cancellation orbit of a given type gives rise to a transverse invariant curve only for a subset of values of the parameter  $\delta$ . This subset has the familiar Cantor

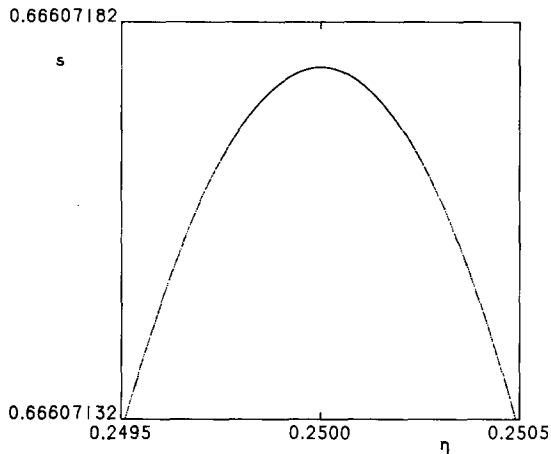


Fig. 16. Enlargement of a part of the cancellation orbit of type a for  $\delta = 0.87637677$ .  $N = 10^6$ .

set structure associated with the KAM theorem: a finite interval is excluded around each  $\delta$  value corresponding to a rational rotation number, i.e. to a cancellation orbit which is periodic.

For a value of  $\delta$  for which a transverse invariant curve does not exist, we may expect the two neighboring chaotic regions to merge into one. An example of this is shown by fig. 17, for  $\delta = 0.8370175289 \dots$ , which corresponds to a periodic cancellation orbit with period 7, rotation

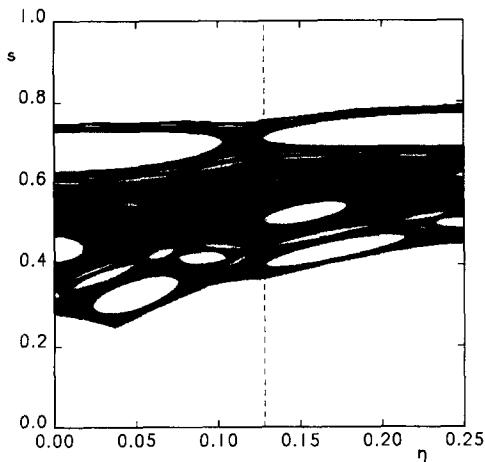


Fig. 17. An orbit which breaks through the barrier of the cancellation orbit a, for  $\delta = 0.8370175289$ . Initial conditions:  $\eta = \eta_2, s = 0.74$ .  $N = 2 \times 10^6$ .

number  $\frac{2}{7}$ . This orbit was started at  $\eta = \eta_2, s = 0.74$ , inside the chaotic region above the cancellation orbit; after about  $10^6$  iterations, it moves into the chaotic region below the cancellation orbit. Fig. 18 shows the evolution of the orbit with time. The abscissa, multiplied by  $10^6$ , gives the number of iterations; the ordinate is  $s$ . Only points lying in the vertical band defined by  $|\eta - \eta_2| < 0.01$  are retained. In each group of 2000 successive iterations, the lowest and highest values of  $s$  are noted and a vertical line joining these values is plotted. Thus, fig. 18 shows the region in which the orbit is wandering, as a function of time. The cancellation orbit corresponds to  $s_a = 0.641850 \dots$ . The orbit of fig. 18 starts above this value, crosses it downwards at about  $N = 10^6$ , and then stays below, except for a short excursion above the barrier in the vicinity of  $N = 1.9 \times 10^6$ .

The two chaotic regions are therefore merged, although  $\delta$  is higher than the critical value 0.76 of Benettin and Strelcyn. The Lyapunov numbers of the two chaotic regions are then replaced by a single number, which is presumably some average between them. Thus, the situation appears to be much more complex than is suggested by fig. 13 in [1]. As  $\delta$  is varied, neighbouring chaotic regions constantly merge and separate again; thus, the number of chaotic regions, the extent of each of

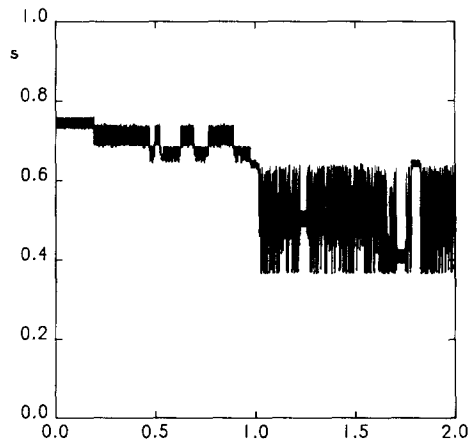


Fig. 18. Variation of  $s$  with time, in the interval  $-0.01 < \eta - \eta_2 < 0.01$ , for the orbit of fig. 17.

these regions, and the values of the associated Lyapunov numbers do not change regularly but rather jump around in highly irregular fashion.

In practice, another phenomenon comes into play. We have found cases where the invariant curve clearly does not exist, yet the passage from one chaotic region to the other was not observed, even after a large number of iterations. A good example of this situation is provided by the case  $\delta = 0.76$  (figs. 5 and 6). In other cases, the diffusion of the points was observed to stop for a long time at the location of the cancellation orbit, until finally they broke through. Apparently, even when there is no true invariant curve, the cancellation orbit can still manifest its presence and act as a barrier, through which passage is not impossible but difficult. In fig. 18, a number of instances of such temporary barriers can be seen. Perhaps the chain of blobs formed by the cancellation orbit acts as a sieve with very small holes; a chaotic orbit then stays on one side until it happens by chance to find one of the holes.

## 6. Conclusions

Our numerical results suggest that mappings with discontinuities in their first derivative, as exemplified by the oval billiard, have peculiar properties which set them apart from the regular case. Invariant curves are generally destroyed when they cross a line of discontinuity; they survive only in particular cases, when a “cancellation of discontinuity” occurs.

It follows that the oval billiard cannot be used as a model for mappings in general, as was too quickly assumed in [8]. On the other hand, it could be used as a prototype for practical problems where discontinuities are present.

## Acknowledgements

This work was partially supported by the North Atlantic Treaty Organization Postdoctoral Fellow-

ship Program administered by the United States National Science Foundation.

## Appendix

Froeschlé [5] studied numerically the mapping

$$\begin{aligned} x' &= x \cos \alpha - (y - |x|) \sin \alpha, \\ y' &= x \sin \alpha + (y - |x|) \cos \alpha, \end{aligned} \quad (\text{A.1})$$

where  $\alpha$  is a parameter. This mapping has a discontinuity in its first derivative across the line  $x = 0$ . In spite of this, Froeschlé found that for many values of  $\alpha$  the plane is filled with invariant curves surrounding the origin, in apparent contradiction with our own results. It turns out, however, that the mapping (A.1) is a very particular case, in that it is “semi-linear”: if  $x$  and  $y$  are multiplied by a positive constant  $\lambda$ ,  $x'$  and  $y'$  are multiplied by the same constant. With polar coordinates defined by  $x = \rho \cos \theta$ ,  $y = \rho \sin \theta$ ,  $\rho \geq 0$ , (A.1) becomes

$$\theta' = f(\theta), \quad \rho' = \rho g(\theta), \quad (\text{A.2})$$

where  $f$  is defined by

$$f(\theta) = \begin{cases} \alpha + \arctan(\tan \theta - 1) & \text{if } -\pi/2 < \theta < \pi/2 \pmod{2\pi}, \\ \alpha + \pi + \arctan(\tan \theta + 1) & \text{if } \pi/2 < \theta < 3\pi/2, \\ \alpha + \theta & \text{if } \theta = \pm \pi/2, \end{cases} \quad (\text{A.3})$$

and  $g$  is related to  $f$  by

$$g^2 \frac{df}{d\theta} = 1, \quad g > 0. \quad (\text{A.4})$$

This last relation reflects the area-preserving property of the mapping.  $f(\theta)$  is of class  $C^1$  but not of class  $C^2$ : its second derivative is discontinuous for  $\theta = \pm \pi/2$ . It is an increasing function of  $\theta$ . Equiva-

tion (A.2a) represents an invertible one-dimensional mapping of a circle onto itself. A number of results have been proved for such mappings (see [9] for a short review). For a given value of the parameter  $\alpha$ , the mapping (A.2a) is characterized by a rotation number  $W$ , which is independent of the starting point  $\theta_0$ . If  $W$  is irrational, the mapping is topologically equivalent to a simple rotation: there exists a function  $\theta(t)$  such that the mapping in the new variable  $t$  reduces to

$$t' = t + 2\pi W. \quad (\text{A.5})$$

From (A.2), (A.4) and (A.5) we derive

$$\rho'^2 \frac{d\theta'}{dt'} = \rho^2 \frac{d\theta}{dt}. \quad (\text{A.6})$$

Therefore the quantity  $\rho^2 d\theta/dt$  is invariant under the mapping, and any curve defined by

$$\rho = K \left( \frac{d\theta}{dt} \right)^{-1/2} \quad (\text{A.7})$$

with  $K$  an arbitrary positive constant, is an invariant curve. This explains the results found in [5].

We have experimented with a mapping which differs from (A.1) by the introduction of a small amount of non-linearity:  $|x|$  is replaced by  $|x| + \epsilon x^2$ . Even for small values of  $\epsilon$ , the invariant curves are destroyed and replaced by chaotic orbits. This confirms that (A.1) is a degenerate case.

## References

- [1] G. Benettin and J.-M. Strelcyn, *Phys. Rev.* A17 (1978) 773.
- [2] G.D. Birkhoff, *Dynamical systems* (American Mathematical Society, Providence, 1927) 173.
- [3] A. Brahic, *Astron. Astrophys.* 12 (1971) 98.
- [4] L.A. Bunimovich, *Funkt. Analiz. Jogo Prilog.* 8 (1974) 73 [*Funct. Anal. Appl.* 8 (1974) 254].
- [5] C. Froeschle, *C.R. Acad. Sci. Paris* A266 (1968) 846.
- [6] C. Froeschle, *C.R. Acad. Sci. Paris* A266 (1968) 747.
- [7] J.M. Greene, *J. Math. Phys.* 20 (1979) 1183.
- [8] M. Hénon, *Proceedings of the Summer School "Chaotic Behaviour in Deterministic Systems"* (Les Houches, July 1981), in press.
- [9] S.J. Shenker, *Physica* 5D (1982) 405.