CIRCULAR BILLIARD

MICHAEL DREXLER *, MARTIN J. GANDER †

Abstract. We analyze the problem of perfectly inelastic billiard on a circular table with exactly one permitted bounce. We present a new and intuitively appealing geometric derivation of the solution. Analyzing the solution with respect to the number of permitted paths for any given scenario, we find an analytical expression for a separatrix between regions with two and four solutions. We identify and discuss symmetry aspects of the problem and singular points on the billiard table. Finally we apply the results to an optical experiment which can be performed in any classroom.

Key words. circular billiard, caustic

AMS subject classifications. 65H05, 78A05

1. Introduction. Imagine a circular billiard table with two billiard balls on it. In which direction does one have to hit the first ball so that it bounces off the cushion once and then hits the other ball? A typical situation is shown in Figure 1.1.
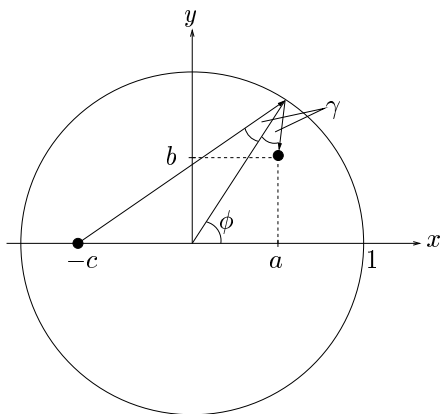

FIG. 1.1. *A typical case of the circular billiard problem*

This problem has been considered in the literature before: Gander and Gruntz [1] solve the problem using the computer algebra system Maple. Bartoň [2] shows two analytical solutions. His first solution relies upon the reflection law which states that the angles between the inbound and outbound track of the moving billiard ball and the normal at the point of impact have to be equal. The second solution uses Fermat's principle that a light ray between two points always chooses the shortest path. Waldvogel [4] derives an analytical solution using complex numbers and Hungerbühler presents an elegant geometrical argument in [3]. Further work on general table shapes has been done in [2]. In the present study, we are interested in a geometrical argument which improves the intuition for finding the number of solutions to this problem. We discuss symmetry aspects and generalize the results to explain an optical experiment via a hitherto unconsidered analogy.

---

*NUMERICAL ANALYSIS GROUP, OXFORD UNIVERSITY, ENGLAND.
†SCIENTIFIC COMPUTING AND COMPUTATIONAL MATHEMATICS, STANFORD UNIVERSITY, USA.

2. Geometric Solution. For an elliptical billiard table with one ball in each focus, any point on the rim is a solution. Hence to solve the billiard problem for the circle, we have to find an ellipse touching the given circle with the two given billiard balls as focal points of the ellipse.

Without loss of generality we choose the coordinate system such that the major axis of the ellipse lies on the x-axis with the focal points symmetric to the origin, as shown in Figure 2.1. The center of the circular billiard table is at (m_1, m_2). To find
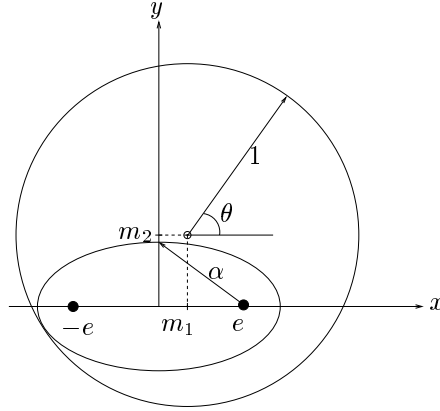

FIG. 2.1. *Coordinate system with the ellipse centered around the origin*

an ellipse touching the circular table tangentially, we have to find a point (x, y) which satisfies the equation of the circle,

$$(2.1) \qquad (x - m_1)^2 + (y - m_2)^2 = 1,$$

the equation of the ellipse,

$$(2.2) \qquad \frac{x^2}{\alpha^2} + \frac{y^2}{\alpha^2 - e^2} = 1$$

and an equation, ensuring that the ellipse touches the circle at (x, y) tangentially. This last condition can be imposed by noting that equations (2.1) and (2.2) are describing level sets of the left hand side functions. Since the gradient of these functions is orthogonal to the level sets, the circle and ellipse touch tangentially at (x, y) if the gradients at that point are parallel. Setting the cross product of the gradients equal to zero, we arrive at the third equation

$$(2.3) \qquad \frac{x}{\alpha^2}(y - m_2) - \frac{y}{\alpha^2 - e^2}(x - m_1) = 0.$$

Before solving this system of nonlinear equations in the variables $\{x, y, \alpha\}$, we would like to gain some intuition from the ellipse model. Suppose the two billiard balls are located close to each other, and away from the boundary. Then a small ellipse with the two balls as its focuses lies completely inside the circle. Increasing the parameter α the ellipse becomes bigger, and also its shape becomes more and more circular, as shown in Figure 2.2 on the left. If the two balls were in the same place, the ellipse would be a circle and thus only touch the other circle twice during the increase of the parameter α. This result remains valid if the balls are close together. If, however,

the balls are far apart, the ellipse touches the circle four times during the increase of the parameter α, as shown in Figure 2.2 on the right. So in general we expect two solutions if the balls are close to each other, and four if they are far apart. But what happens between these limiting cases? To analyze this, we solve the system (2.1), (2.2), (2.3). Prior to attempting a general solution, we consider two pathological cases, which have to be excluded from the general solution.

For $e = 0$, the problem reduces geometrically to finding the points where two circles with displacement vector $\mathbf{d} = (m_1, m_2)^T$ touch. In this context, one circle (the billiard table) is the unit circle and the other circle has radius α. There are two solutions on the line defined by $\mathbf{d}$, with

$$\alpha_1 = 1 - \|\mathbf{d}\|, \quad \alpha_2 = 1 + \|\mathbf{d}\|,$$

where $\| \cdot \|$ denotes the Euclidean norm.

For $m_1 = m_2 = 0$, the ellipse is centered with the circle. We have thus four solutions, two at $y = 0$, $x = \pm 1$ and two at $x = 0$, $y = \pm 1$. The parameter α for these cases is found to be

$$\alpha_1 = 1, \quad \alpha_2 = \sqrt{1 + e^2}.$$

For a general solution, we parametrize the circle by $x := m_1 + cos(\theta)$ and $y := m_2 + sin(\theta)$, solve equation (2.3) for α^2,

$$(2.4) \qquad \alpha^2 = \frac{xe^2(m_2 - y)}{ym_1 - xm_2}$$

and substitute the result together with the parametrized coordinates into equation (2.2). This equation depends now only on θ. To eliminate the trigonometric functions, we introduce the transformation $\theta = 2\arctan(u)$. Equation (2.2) becomes

$$(2.5) \qquad \frac{Q(u)}{e^2 u(u-1)(u+1)} = 0,$$

where

$$(2.6) \qquad \begin{aligned} Q(u) = \ & (m_2 - m_2 m_1)u^4 + (2m_1 - 2m_1^2 + 2e^2 + 2m_2^2)u^3 \\ & + 6u^2 m_2 m_1 + (-2m_2^2 + 2m_1^2 + 2m_1 - 2e^2)u - m_2 m_1 - m_2. \end{aligned}$$

To find a solution u from (2.5), we can ignore the denominator except if the solution is $u = 0$ or $u = \pm 1$. We show now that we can ignore the denominator in these cases as well:

$u = 0$ implies $\theta = 0$ and thus the y-coordinate of this solution is $y = m_2$. Since at $y = m_2$ the circle has a vertical tangent, the only possibility for a solution at that point is that the ellipse has a vertical tangent as well, because at the solution the ellipse touches the circle tangentially. Hence the center of the circle must lie on the big axis of the ellipse and therefore $y = m_2 = 0$. In that case our derivation breaks down, since the denominator in (2.4) vanishes. Nevertheless the numerator provides the appropriate solution, since $Q(0) = 0$ for $m_2 = 0$.

$u = \pm 1$ implies $\theta = \pm\frac{\pi}{2}$ and therefore the x-coordinate of this solution is $x = m_1$. Note that at $x = m_1$ the circle has a horizontal tangent, so the only possibility for a solution at that point is that the ellipse has a horizontal tangent as well. Thus the center of the circle must lie on the small axis of the ellipse and therefore $x = m_1 = 0$.

Again our derivation breaks down, since the denominator in (2.4) vanishes. But as in the previous case, the numerator provides the appropriate solution, since $Q(\pm 1) = 0$ for $m_1 = 0$.

Therefore it is sufficient to analyze the fourth order polynomial equation

$$(2.7) \qquad\qquad Q(u) = 0$$

to determine the solutions to our problem. If the billiard balls are close together, equation (2.7) has two real and two complex roots, whereas if the balls are far apart, it has four real roots, as shown in Figure 2.2.
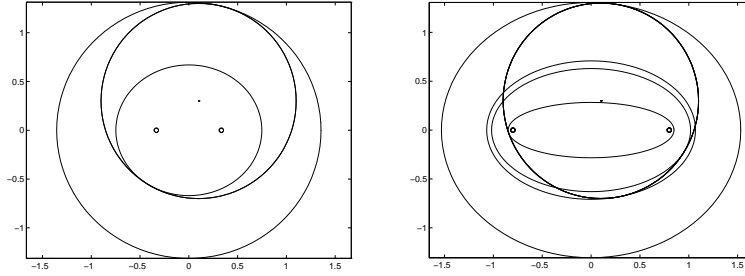

FIG. 2.2. *A case with two and one with four solutions*

3. Number of Valid Solutions. To study the number of solutions dependent on the position of the balls, it is more convenient to transform the coordinates so that the circular billiard table has its center in the origin, as given in Figure 1.1. Transforming the polynomial (2.6) accordingly, we obtain

$$(3.1) \quad P(v) := b(c+1)v^4 + 2(c+2ac+a)v^3 - 6bcv^2 + 2(c-2ac+a)v + b(c-1) = 0.$$

where we used the notation in Figure 1.1, namely
- The first ball is positioned at the Cartesian coordinates (a, b), with the constraint $a^2 + b^2 \leq 1$, on a billiard table of unit radius.
- The second ball is located at the position $(c, 0)$, $-1 < c \leq 0$.
- The parameter v is related to ϕ by the substitution $\phi = 2\arctan(v)$.

It is noted that any other scenario on the billiard table can be obtained from this via a simple rotation. We emphasize that equation (3.1) is identical with the equation derived in [2] using the law of reflection.

We first conduct a computer experiment to demonstrate the problem. Fixing one ball on the x-axis at the position $(c, 0)$, we let the other ball at (a, b) vary and compute the number of real roots we obtain from (3.1). Doing this on a fine mesh, we see the interesting result that there are only two connected regions on the circular billiard table, one (light grey) which allows for four solutions and one (dark grey) which allows for two. An example is shown in Figure 3.1. To compute the separatrix of those two regions, note that two roots of the polynomial $P(t)$ are becoming complex if a zero of the derivative of $P(t)$ moves through the x-axis. Thus to find the separatrix as a function of t for given c in parameter form, we have to solve the system of equations

$$P(t) = 0$$
$$P'(t) = 0$$

FIG. 3.1. *Regions on the billiard table with two and four solutions, $c = -0.9$*

for the position (a, b). Note that $P(t)$ is linear in a and b, so that the above system is linear and can be solved analytically. The solutions are given by

$$(3.2) \quad \begin{aligned} a_0 &= -\frac{c}{h} \left[(1 + c)t^6 + 3(1 + 3c)t^4 + 3(1 - 3c)t^2 + (1 - c) \right], \\ b_0 &= \frac{16}{h} c^2 t^3, \\ h &= (1 + 3c + 2c^2)t^6 + 3(1 + c + 2c^2)t^4 + 3(1 - c + 2c^2)t^2 + (1 - 3c + 2c^2). \end{aligned}$$

The connected set of points (a_0, b_0) for $t \in (-\infty, \infty)$ defines the separatrix between the regions with two and four solutions. In this notation, t serves as a parameter to identify a point (a_0, b_0) on the curve. Due to the symmetry arguments outlined in the next section, there can be no less than two solutions for $b \neq 0$. The discussion of singular (i.e. non-differentiable) points on the separatrix will be postponed until then.

Figure 3.2 shows the separatrix for different values of the fixed-ball parameter c. The locus of the region with two solutions is evident from comparison with Figure 3.1. To determine the number of solutions on the separatrix, we substitute (3.2) into (3.1) to obtain

$$(3.3) \quad \begin{aligned} P_s(t, v) &= \frac{8c^2}{h_s}(v - t)^2 \left[2t^3(1 + c)v^2 + \left(t^4(1 + c) + 6ct^2 + c - 1 \right) v + 2t(c - 1) \right], \\ h_s &= \left(1 + t^2 \right)^2 \left[(2c^2 + 3c + 1)t^2 + 2c^2 - 3c + 1 \right]. \end{aligned}$$

This determines a polynomial in v for each point defined by t on the separatrix. The number of roots of that polynomial determines the number of solutions at that point. We note that P_s has a double root at $v = t$; therefore, the number of solutions can at most be 3. The determinant of the second quadratic term of P_s in (3.3) shows that no other double root can exist for $|c| \leq 1$. To examine the existence of triple roots, we differentiate twice with respect to v and get at the point of interest $v = t$

$$P_s''(t, t) = \frac{48c^2 t}{h_s} \left[(1 + c)t^4 + 2ct^2 + c - 1 \right]$$

with h_s as defined in (3.3). This yields three points on the separatrix where P_s has a triple root at $v = t$,

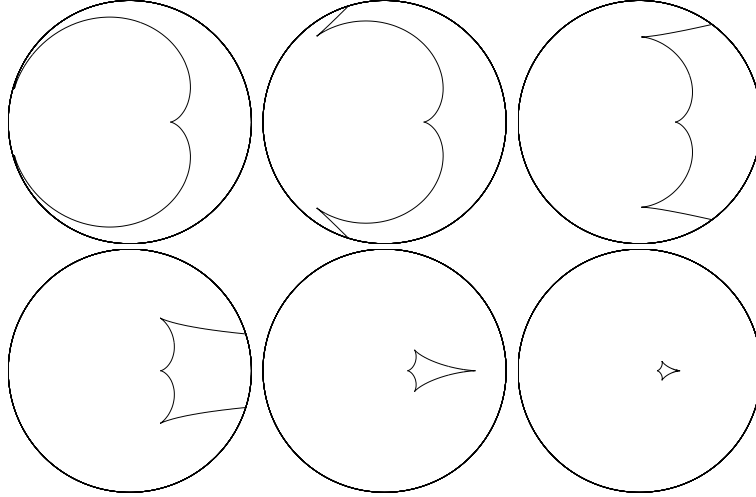$$t_0 = 0, \quad t_{1,2} = \pm\sqrt{\frac{1 - c}{1 + c}}.$$

Figure 3.3 depicts the change in the polynomial as $t \geq 0$ marches along the separatrix. For the chosen $c = -0.6$, the critical value is $t_1 = 2.0$. The pictures for negative t are similar with the direction of the axes reversed.

Starting from the degeneracy at $t = 0$, the double root is bracketed by the two simple roots for $|t| < |t_{1,2}|$ and marches through the positive$(t > 0)$ or negative$(t < 0)$ simple root at the point of degeneracy. For larger $|t|$, it resides outside the two simple roots. Of course, this solution in v can be translated into a geometrically relevant angle ϕ by using the substitution proposed earlier.

It can thus be said that in general, two of the four solutions merge into one on the separatrix. This double solution then disappears into the complex plane as the separatrix is left for the region with two solutions. It is easy from Figure 3.3 to visualize this process via vertical shifting of the polynomial P_s.

4. Singularity and Symmetry. For $a = b = c = 0$, the polynomial P degenerates to $P(v) \equiv 0$. This is consistent with the geometrical view of the problem. Every straight line starting from the origin is normal to the circle and therefore reflected back into the origin. It is noted that the singularity at the origin arises from a smooth variation of the polynomial coefficients and thus the polynomials become closer and closer to a zero function, numerically harder to solve, but never does the number of solutions become larger than 4. Only at the origin does the number of solutions jump from 4 to ∞. As $(a, b) = (c, 0)$ implies the two balls sitting on top of each other, it is of course a physically impossible setting.

On the x-axis with $b = 0$, $P(v)$ degenerates to a cubic. The process of degeneration, as indicated in Figure 3.3, is such that one root of the quartic converges closer to zero, and one root gets larger and larger, reaching ∞ in the limit $b \to 0$. This is in accordance with the angle $\phi = \pi$ not being allowed in the trigonometric substitution for v stated earlier - by introducing this unphysical solution (one ball would have to pass the other after hitting), the quartic character could be retained. Due to the cubic character, there is at least one real solution for $b = 0$ with $\phi = 0$. Exactly two solutions are forbidden due to symmetry considerations - any reflection point with
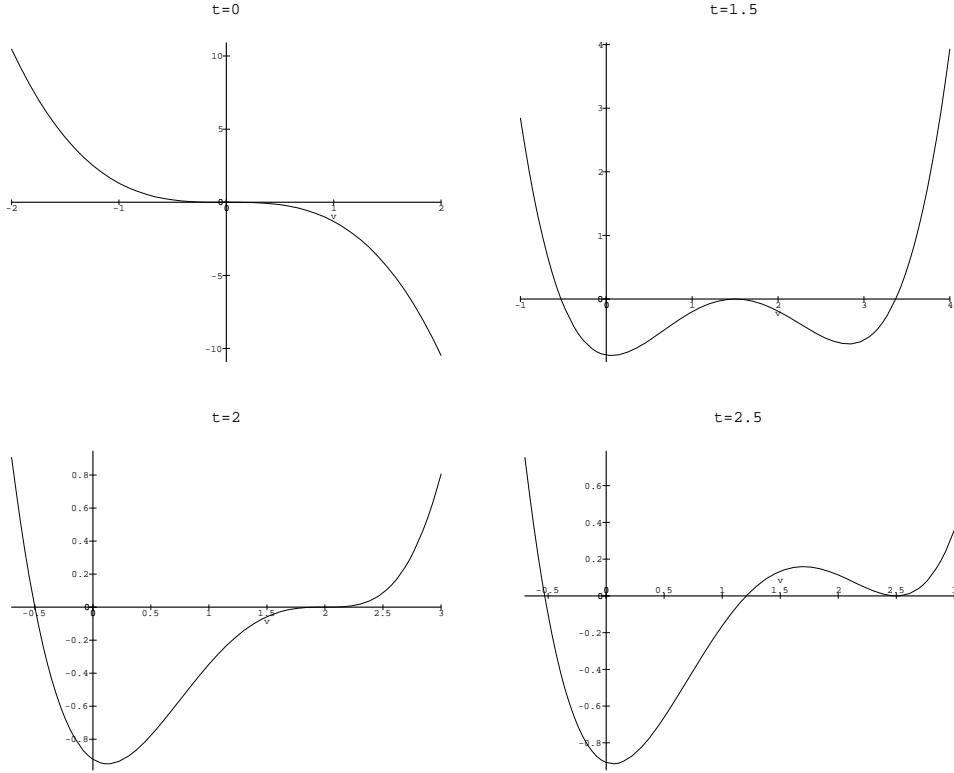
FIG. 3.3. P_s for $t = 0, 1.5, 2.0, 2.5$ along the separatrix for $c = -0.6$

$\phi \neq 0$ must have a symmetric point in the other half-circle that would also constitute a valid solution. Therefore, excluding the solution $\phi = \pi$, the x-axis contains either one or three solutions to the billiard problem.

The solution pattern should by symmetry be invariant with respect to reflections about the x-axis. The separatrix as noted in (3.2) does obviously exhibit this symmetry with $a_0(-t) = a_0(t)$ and $b_0(-t) = -b_0(t)$.

For $b > 0$, we can determine

$$\lim_{v \to \pm\infty} P(v) = +\infty, \quad P(0) = -b(1 - c) < 0.$$

Therefore, as $P(v)$ is of order 4 for non-vanishing b, at least two real roots must exist.

As there are two regions with either two or four solutions to the billiard problem, the only set of points that needs to be analyzed further is the separatrix of these regions. Since the number of roots on this curve has been previously established, we are here concerned with singularities of the curve itself. To find the singular points on the parametrized separatrix (3.2), we compute the derivatives of a_0 and b_0 with respect to t. Setting them to zero simultaneously yields the condition

$$\begin{aligned}
24c^2 t \quad [c - 1 + t^2(1 + c)] \quad [c - 1 - 6ct^2 - t^4(1 + c)] &= 0, \\
48c^2 t^2 \quad [c - 1 + t^2(1 + c)] \quad [2c - 1 - t^2(1 + 2c)] &= 0,
\end{aligned}$$

7

which has three solutions for general c,

$$t_0 = 0, \quad t_{1,2} = \pm\sqrt{\frac{1-c}{1+c}}.$$

From these, we can obtain a first approximation to the appearance of the separatrix for given c. The position of the first singular point on the x-axis can be obtained by substituting $t = 0$ in (3.2). It is determined by

$$a_0(t_0) = -\frac{c}{1-2c}.$$

The singular points for non-vanishing t are obtained by

$$a_0(t_{1,2}) = c\left(2c^2 - 1\right),$$
$$b_0(t_{1,2}) = \pm 2c^2(1+c)\sqrt{\frac{1-c}{1+c}}.$$

We note that for $|c| < \frac{1}{3}$, there is a fourth singular point inside the unit circle on the x-axis into which the two branches of the separatrix converge for $|t| \to \infty$. The position of that point is governed by

$$a_0(t \to \infty) = -\frac{c}{1+2c}.$$

It is noted that the singular (i.e. non-differentiable) points of the separatrix are also these points where the number of solutions to the billiard problem changes from two to four (one to three on $b = 0$) abruptly via a triple zero of P_s without the intermediate three solutions generally present on the separatrix.

5. An Experiment. In this section, we want to show how the separatrix can be seen in a simple optical experiment. Imagine a point-like (or very small) light
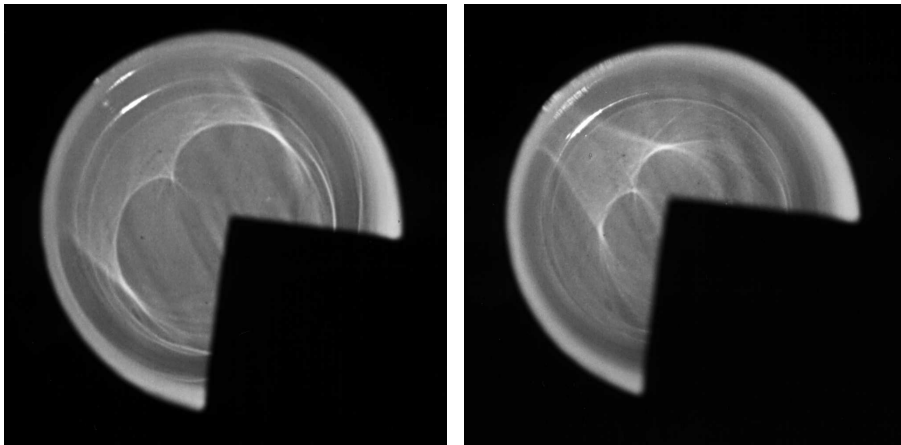

FIG. 5.1. *The optical caustic for two positions of the light source*

source being positioned in a cylindrical geometry with a flat bottom. For an easy experiment, an empty coffee mug and a flashlight bulb will do. In the projection onto the bottom, the light source can be identified with the second billiard ball at $(c, 0)$.

8

Our eyes run the computer experiment outlined earlier in an instantaneous, parallel fashion, with the path being reversed. At each point (a, b) on the circular 'table', how many reflected light rays arrive from the source? This determines the intensity of that point to our eyes. As expected, the brighter region is the one with four solutions, and the darker one has only two solutions, similar to the pattern in Figure 3.1. The special case $b = 0$ is remedied with the now physical inclusion of $\phi = \pi$ as a solution. The separatrix is clearly visible and as we vary the position of the light source, its shape changes just like depicted in Figure 3.2. Two positions of the point light source are photographed in Figure 5.1. The light source has been covered with a wedged blind to avoid over exposure.

6. Conclusion. We treated the circular billiard problem in a comprehensive fashion, outlining both geometric and analytic aspects of the solution. We depicted regions with a given number of solutions and derived their separatrix in analytic form. We identified and discussed degenerate points on the billiard table. Relating the results to an optical experiment, which can be easily performed in a classroom or at home, we made the material more accessible.

REFERENCES

[1] W. Gander and D. Gruntz, *The Billiard Problem*, Int. J. Educ. Sci. Technol., 23 (1992), pp. 825-830.
[2] S. Bartoň, *The generalized Billiard Problem*, in: W. Gander and J. Hřebíček, *Solving Problems in Scientific computing using Maple and MATLAB*, Springer-Verlag Berlin Heidelberg 1993, 1995.
[3] N. Hungerbühler, *Geometrical Aspects of the Circular Billiard Problem*, El. Math., 47 (1992), pp. 114-116.
[4] J. Waldvogel, *The Problem of the Circular Billiard*, El. Math., 47 (1992), pp. 108-113.
[5] G.B. Whitham, *Linear and Nonlinear Waves*, Wiley, New York (1974).