Periodic trajectories in right-triangle billiards

Barry Cipra,* Robert M. Hanson,† and Amy Kolan‡
Saint Olaf College, 1520 Saint Olaf Avenue, Northfield, Minnesota 55057-1098
(Received 8 March 1995)

Billiard problems are simple examples of Hamiltonian dynamical systems. These problems have been used as model systems to study the link betwen classical and quantum chaos. The heart of this linkage is provided by the periodic orbits in the classical system. In this article we will show that for an arbitrary right triangle, almost all trajectories that begin perpendicular to a side are periodic, that is, the set of points on the sides of a right triangle from which nonperiodic (perpendicular) trajectories begin is a set of measure zero. Our proof incorporates the previous result for rational right triangles (where the angles are rational multiples of π), while extending the result to nonrational right triangles.

PACS number(s): 03.20.+i

The game of billiards provides a profusion of interesting questions in classical mechanics. Coriolis [1] and Sommerfeld [2] were intrigued by questions concerning high and low shots, the causes and effects of "English," and the beautiful curved paths resulting from friction with the billiard cloth.

Recently there has been a resurgence of interest in billiard problems, but of quite a different type. In this new type of problem, the billiard balls are point particles, which travel in straight lines and bounce elastically off the sides of the table. The table, though, has lost its familiar rectangular shape; it may now be a more general polygon, ellipse, or stadium (a rectangle capped by two semicircles). This species of billiards is a particularly simple example of a chaotic dynamical system. As such, it has been used as a model system to study the link between classical and quantum chaos [3].

The heart of this linkage is provided by the periodic orbits in the classical system, which provide information on the density of states in the quantum system. In addition, the beguiling simplicity of these periodic orbits lends them an intrinsic mathematical appeal. The study of periodic orbits in a simple polygon or other geometric shape is, after all, a problem that Euclid could have pondered.

Consider, for example, a trajectory on a right-triangular billiard table that starts out at right angles to one of the two sides of the right angle. After 14 bounces, the trajectory shown in Fig. 1 hits the same side again at right angles, at which point it simply retraces its path. The trajectory is periodic.

For right triangles whose other angles are rational multiples of π it is known that all trajectories that start out at right angles to one of the two sides are periodic, except for those that hit one of the vertices of the triangle [4,5].

We shall show that for an arbitrary right triangle, *almost all* trajectories that begin perpendicular to a side are periodic, that is, the set of points on the sides of a right triangle from which nonperiodic (perpendicular) trajectories begin is a set of measure zero. Our proof will also give the previously known result for rational right triangles.

The problem of right-triangle trajectories is intimately related to another problem, that of two elastic point particles colliding with each other in the unit interval [6]. The point particles have masses m_1 and m_2, and there are walls at 0 and 1. In mapping from one problem to the other, the positions of the two particles in the interval, weighted by the square roots of their masses, correspond to the x and y coordinates of a point in a right triangle where one angle is $\arctan(m_1/m_2)$.

We begin with the standard representation of a trajectory as a sequence of flips of the triangle along a straight line [6,7], as shown in Fig. 2. Notice that the same sequence of flips applies to all trajectories in a band about the chosen trajectory. The upper and lower limits of the band are determined by where the vertices of the flipped triangles land: The upper limit is determined by the lowest vertex above the chosen trajectory, the lower limit by the highest vertex below.

Notice also that one of these vertices (in this case the one above) is the right-angle vertex. If we look at a trajectory which is as much above this vertex as the original

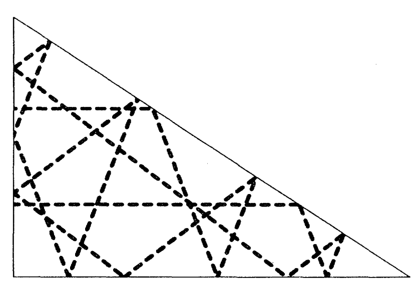

FIG. 1. A fifteen-segment periodic trajectory within a right triangle starting perpendicular to the short leg. In this case the ratio of the short to long leg is 1:1.42.

*Electronic address: cipra@stolaf.edu
†Electronic address: hansonr@stolaf.edu
‡Electronic address: kolan@stolaf.edu

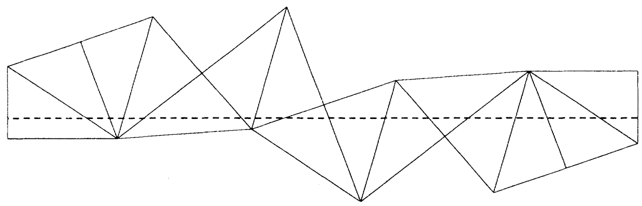

FIG. 2. A representation of the trajectory of Fig. 1 as a sequence of flips of the triangle.

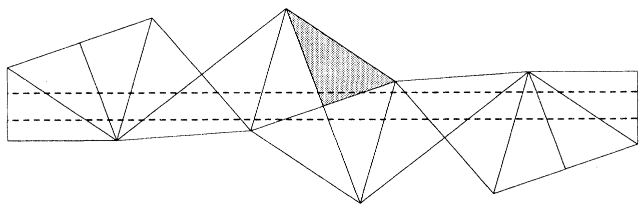

FIG. 3. The trajectory of Fig. 1 along with its equivalent around the point of symmetry, suggesting a rhombus representation.

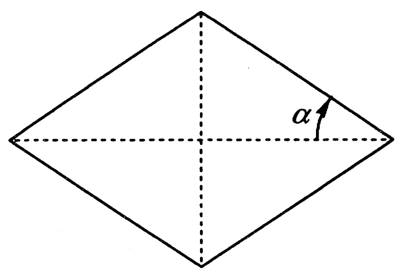

FIG. 4. The rhombus corresponding to the triangle of Figs. 1–3. $\tan\alpha = 1/1.42$.

trajectory was below it, we see that the sequence of flips is essentially the same (see Fig. 3). In fact, as billiard trajectories inside the triangle, *these two are identical.* The right-angle vertex is a point of rotational symmetry in Fig. 3. (We shall comment further on this near the end of the paper.)

This suggests that the right-angle vertex is a "removable singularity." We can remove it by using as our basic shape not the right triangle, but a rhombus consisting of four copies of the triangle, as shown in Fig. 4. We shall henceforth work with rhombi, and consider trajectories which begin perpendicular to one of the diagonals of the rhombus.

As we follow the flips of the rhombus along a straight-line trajectory, we see that each flip in effect rotates the rhombus by an amount equal to one of the two interior angles of the rhombus. It is convenient to describe the rotations in terms of one angle only. (A clockwise rotation by one interior angle is equivalent to a counterclockwise rotation by the other.) If α is one of the angles of the right triangle, then as the rhombus flips along a trajectory, its orientation increases or decreases by 2α with each flip. We can thus label the rhombi with integers, according to the total number of, say clockwise, rotations by 2α for each, as shown in Fig. 5.

This suggests a new representation of a trajectory, as a sequence of line segments in a set of labeled rhombi. An example is shown in Fig. 6. The segments of the trajectory are numbered as they enter each rhombus from the left. When a segment exits a rhombus on the right, it either advances to the next rhombus, or goes back to the previous one, reentering at the corresponding point on the edge parallel to the edge it has just left.

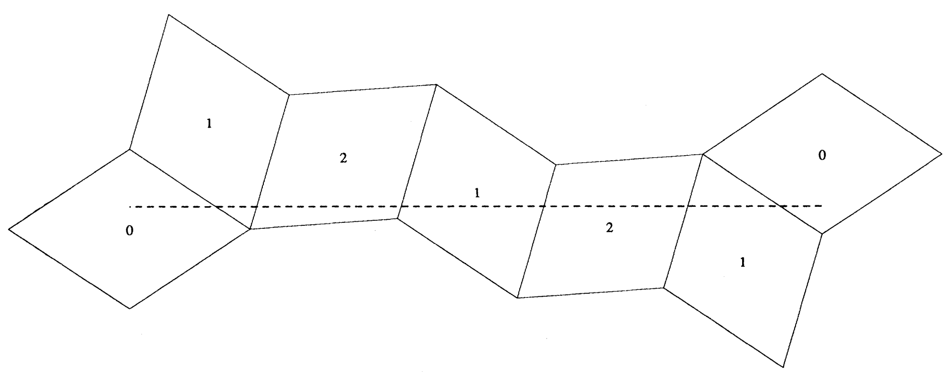

FIG. 5. The trajectory of Fig. 1 as a sequence of rhombus flips. Numbers 0–2 refer to unique rhombus orientations.

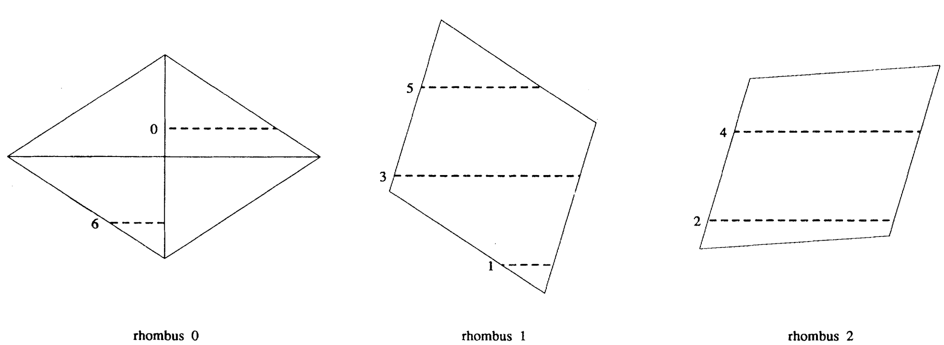

FIG. 6. The three unique rhombus orientations of Fig. 5 with sequential line segments of the trajectory indicated.

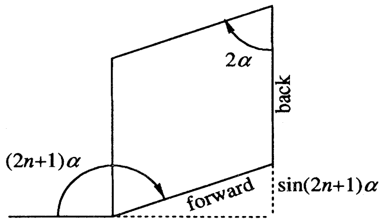

FIG. 7. A generic "nth" rhombus.

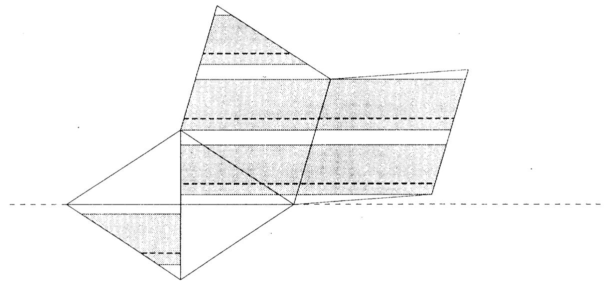

FIG. 9. The rhombi of Fig. 8 arranged in "pinwheel" configuration.

Figure 7 shows a generic "nth" rhombus. The two sides through which a horizontal, right-moving trajectory can exit are labeled "forward" and "back," according to whether the trajectory is to go forward to the next rhombus or back to the previous one. In the convention we have adopted, the forward edge is always the leading (clockwise) ray for the angle 2α, and the backward edge is the trailing (counterclockwise) ray. (It may help to point out that when the trajectory goes forward, it enters the next rhombus on the unlabeled edge adjacent to that rhombus's forward edge, and likewise, when it goes backwards, it enters on the unlabeled edge adjacent to the backward edge.)

When α is a rational multiple of π, there are only finitely many distinct orientations for the labeled rhombi, while for "irrational" angles α, there are infinitely many orientations. Note that the forward edge in Fig. 7 is nearly horizontal. Since the forward edge of rhombus 0 is set at angle α with respect to the x axis, and each subsequent forward edge is a clockwise rotation by 2α, the vertical width of the forward edge of rhombus n is $|\sin[(2n+1)\alpha]|$ (we take the rhombus to have sides of unit length). An elementary result in number theory guarantees that when α is an irrational multiple of π, $|\sin[(2n+1)\alpha]|$ is never equal to 0, but takes values that are arbitrarily close to 0 as n ranges over the positive integers. (The number-theoretic result is that the set of fractional parts of the integer multiples of any irrational number is dense in the unit interval [8].) The fact that the vertical width of the forward edge of the nth rhombus is arbitrarily small for certain values of n is one of the keys to proving that almost all trajectories are periodic.

Recall now that each periodic trajectory actually belongs to a band of trajectories that follow the same pattern of flips. This carries over to the new representation, as shown in Fig. 8. The width of the band is determined by the vertices P_1 and P_2, which separate the forward and backward edges of rhombus 1 and rhombus 2, respectively. It is clear that any periodic trajectory can be widened into such a band.

A more compact, picturesque form of Fig. 8 is shown in Fig. 9. This "pinwheel" configuration of the rhombi makes it easier to see the progression of the band. However, the analysis is based on considering each rhombus separately.

Note that the width of the band remains constant. This is clear, since the rule that ties trajectories from one rhombus to another simply amounts to an affine translation.

Note also that the segments of the band, such as segments 3 and 5 in rhombus 1, *do not overlap.* This is easy to prove: If segments h and k overlapped in one rhombus, with $h<k$, then segments $h-1$ and $k-1$ would have overlapped in the preceding rhombus. (This argument would fail if h were 0, since there is no "-1" segment, but since segment 0 starts in rhombus 0, which is the one that is symmetrically oriented, there can be no segment that overlaps with it: Any segment that reenters rhombus 0 does so in its lower half, and stops at right angles at the diagonal from which segment 0 began.)

This nonoverlapping of segments holds between bands as well, for the same reason. That is, let B_1 and B_2 be two bands of trajectories, each starting at right angles to the vertical diagonal of rhombus 0. Then no segment of band B_1 overlaps with any segment of band B_2.

The nonoverlapping of bands and the constancy of their width are two more keys to proving that almost all

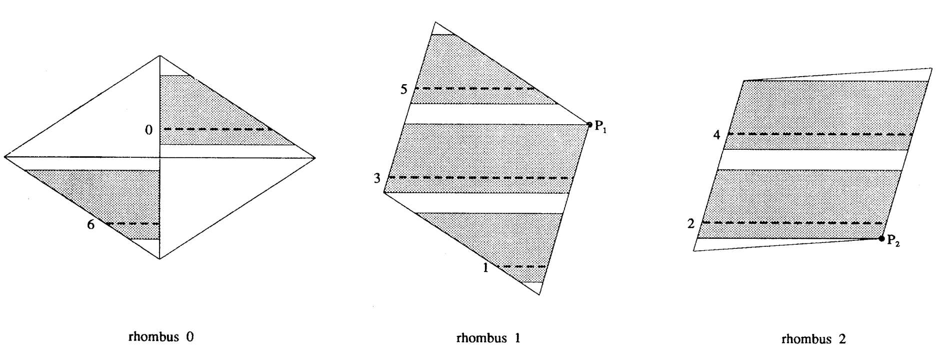

FIG. 8. The band of trajectories containing the trajectory of Fig. 6 and bounded by points P_1 and P_2.

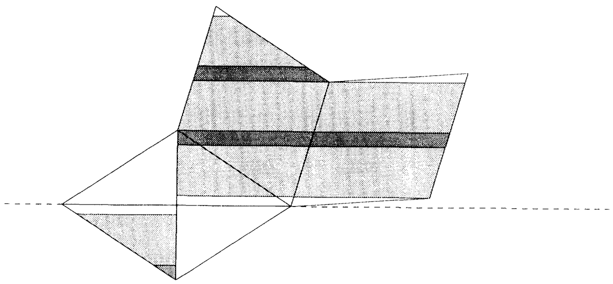

FIG. 10. The two bands of trajectories involving only the first three rhombus orientations and accounting for approximately 87% of all trajectories beginning perpendicular to the short leg.

trajectories are periodic. We are now in position to prove our main result.

Consider what happens to a "beam" consisting of all trajectories perpendicular to the vertical diagonal in the top half of rhombus 0. It is transmitted as a whole to rhombus 1, as shown in Fig. 10. In that example, the beam is again transmitted as a whole to rhombus 2, but that rhombus *splits* the beam, sending the lower part of it on to rhombus 3, and the higher part back to rhombus 1. We observe that the latter subbeam is "trapped" in rhombi 1, 2, and 0: The only entrance to rhombus 3 is through the forward edge of rhombus 2, but that edge is fully filled, and beams, like bands, cannot overlap. The trapped subbeam is split again, this time by the vertex in rhombus 1, but both of these subbeams finally wind up in rhombus 0, where they terminate.

The situation just described, in which the top part of the initial beam is trapped in the first few rhombi while the bottom part goes further, happens whenever there is an integer m (depending on α) for which $0<\sin[(2m+1)\alpha]<\sin(\alpha)$ but $\sin(\alpha)<\sin[(2k+1)\alpha]$ for $0<k<m$: The initial beam advances as a whole to rhombus m, where it is split [9]. The bottom, "transmitted" subbeam fills up the forward edge leading to rhombus $m+1$, while the top subbeam is trapped in rhombi $0-m$. The trapped subbeam may be split again by the vertices of rhombi $1,2,\ldots,m-1$, but all subbeams eventually wind up in rhombus 0, where they terminate. Why? *Because beams of constant width cannot continue indefinitely in a finite region without overlapping.*

For emphasis, we note that, due to the nonoverlapping condition, each vertex separating the forward and backward edges of a rhombus can act as a "beam splitter" at most once. Thus the "trapped" beam we have just been describing splits (if at all) into at most m subbeams.

Figure 11 shows the continuation of the transmitted beams in rhombi 3, 4, and 5. Only a very thin sliver (less than 2.5%) of the original beam is sent on to rhombus 6; the rest, consisting of three subbeams (two "trapped" beams and most of the "transmitted" beam), is confined to rhombi 0–5. Figure 12 shows the same picture in the pinwheel configuration.

As noted above, if a right triangle has angles that are rational multiples of π, then there are only a finite number of rhombi. Therefore, the beam, which is split at most a finite number of times, cannot contain a subbeam that continues indefinitely. This suffices to prove the known result, that all trajectories perpendicular to a side (except those that hit a vertex) are periodic.

But what about the irrational case? In general, when a beam enters rhombus n, one of three things happens: It can exit entirely through the forward edge and go on to rhombus $n+1$, it can exit entirely through the backward edge and be sent back to rhombus $n-1$, or it can be split in two. But, as we established earlier, there are integers n for which the vertical width of the forward edge of the nth rhombus is arbitrarily small. Consequently, for such an n, an arbitrarily large fraction of the initial beam will be trapped in rhombi $0,1,..,n$, in the form of at most n

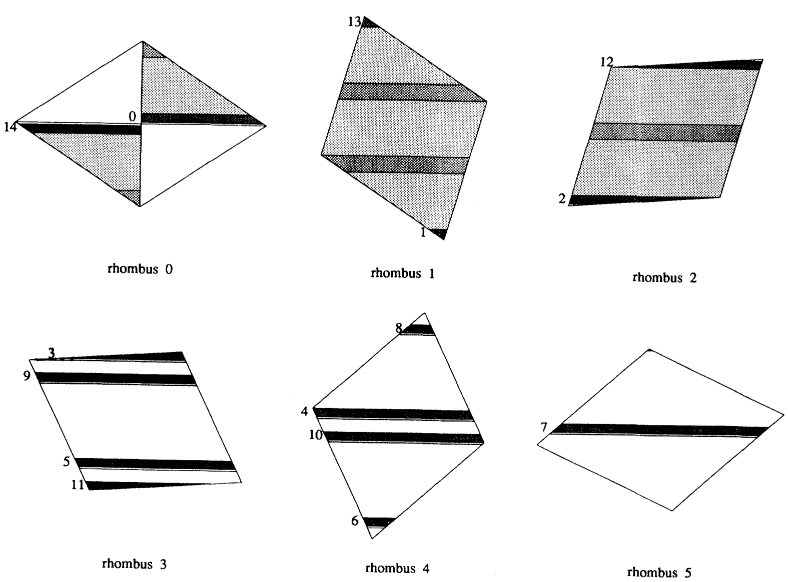

FIG. 11. The three bands of trajectories involving only the first six rhombus orientations and accounting for approximately 97.7% of all trajectories beginning perpendicular to the short leg.

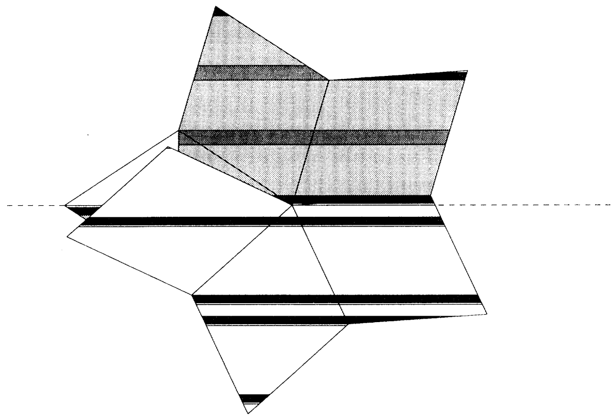

FIG. 12. The pinwheel configuration of Fig. 11.

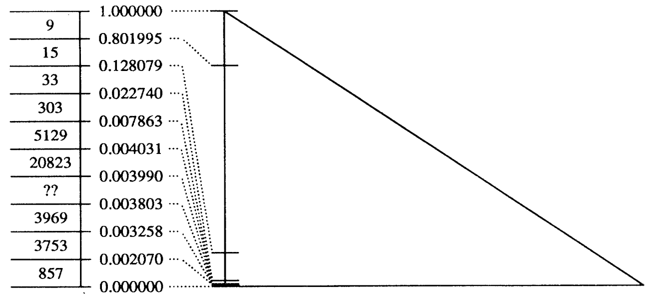

FIG. 13. Approximate boundaries of all bands of trajectories that we have been able to identify in the triangle of Fig. 1, accounting for 99.98% of all trajectories beginning perpendicular to the short leg. On the left is given the number of line segments composing the trajectories of each of the nine identified bands.

subbeams. The trajectories comprising these beams (excluding those that hit vertices) are necessarily periodic. The nonperiodic trajectories, if any, are contained in beams of increasingly narrow width, and hence constitute a set of measure zero. The result is proved: Almost all trajectories perpendicular to a side of a right triangle with irrational angles are periodic.

Our analysis leaves open several questions worthy of further study. First and foremost, are there really any nonperiodic trajectories emanating at right angles from the side of an arbitrary right triangle, and if so, how many? Is the number finite, countable, or uncountable (yet still, like a Cantor set, of measure zero)?

Computational evidence suggests the number may be either 0 or 1. Figure 13, for example, shows what we know about the trajectories in the triangle of Figs. 1–12, which we take to have height 1 and base 1.42. The decimal numbers on the left identify the upper and lower boundaries of the beams; the integers specify the number of segments constituting the trajectory in the triangle. Note that there is a "mystery" interval between 0.003 803 and 0.003 990, where the number of segments is undetermined. This represents the limit of the computation: Almost all of the trajectories in this beam are periodic, but their periods exceed our program's ability to compute the trajectories with sufficient accuracy.

The fact that all triangles we have looked at to date have only one such "mystery" interval is what inclines us to speculate that there may be at most one nonperiodic trajectory left over. However, more detailed computation may well show the mystery interval splitting into myriad mystery intervals separated by well-behaved beams of periodic trajectories. A more detailed analysis of the way beams are split by rhombi may also resolve the question one way or the other.

One positive feature of nonperiodic trajectories is worth noting: Our proof shows that any such trajectory is a limit of periodic trajectories of longer and longer period.

Our computational evidence also suggests that the trajectory at the middle of each beam hits the right-angle vertex of the triangle. This recalls the situation depicted in Fig. 3. When this happens, the trajectories above and below the midpoint are identical: A trajectory in the beam that leaves the side of the triangle at a distance d above the midpoint returns at distance d below, and vice versa. It is clear that every trajectory that hits the right-angle vertex is at the middle of a beam of periodic trajectories, but it is not clear (to us, at least) that every periodic beam has such a trajectory at its middle.

Finally, we note that our proof applies equally well to trajectories that start at right angles to the base of an isosceles triangle, but not to trajectories in a nonright, scalene triangle. Indeed, it remains an open question whether an arbitrary triangle with an obtuse angle necessarily has any periodic trajectories at all [7]. Euclid left us a lot of unfinished business.

We would like to thank Eugene Gutkin, Serge Troubetzkoy, and Serge Tabachnikov for their advice, comments and encouragement. Thanks also to Loren Larson, Jim Cederberg, Duane Olson, William Thurston, John Conway, Ron Graham, Leo Kadanoff, Sid Nagel, and Tom Witten.

[1] Gaspard Coriolis, *Théorie Mathematique des Effects du jeu de Billiard* (Carilian-Goeury, Paris, 1835).

[2] Arnold Sommerfeld, *Mechanics: Lectures on Theoretical Physics* (Academic, New York, 1952), Vol. 1, pp. 158–161.

[3] Martin Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer-Verlag, New York, 1990).

[4] G. Galperin and A. Zemlyakov, *Mathematical Billiards* (Nauka, Moscow, 1990).

[5] Michael Boshernitzan, Am. Math. Mon. **99**, 522 (1992).

[6] Eugene Gutkin, Physica D **19**, 311 (1986).

[7] Ya. B. Vorobets, G. A. Galperin, and A. M. Stepin, Russ. Math. Sur. **47**, 5 (1992).

[8] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford University Press, Oxford, 1979), p. 375. The number-theoretic result is known as Kronecker's Theorem.

[9] The condition is equivalent to the double inequality $0 \le \pi - (2m+1)\alpha < \alpha$; such an integer m exists if and only if the integer part of π/α is odd.