ANZIAM J. **67, e33** (2025), 1–30
doi:10.1017/S1446181125100163

CLASSIFICATION OF PERIODIC ORBITS FOR SQUARE AND RECTANGULAR BILLIARDS

HONGJIA H. CHEN[1] and HINKE M. OSINGA[1]

(Received 26 June, 2025; accepted 2 July, 2025)

Abstract

Mathematical billiards is much like the real game: a point mass, representing the ball, rolls in a straight line on a (perfectly friction-less) table, striking the sides according to the law of reflection. A billiard trajectory is then completely characterized by the number of elastic collisions. The rules of mathematical billiards may be simple, but the possible behaviours of billiard trajectories are endless. In fact, several fundamental theory questions in mathematics can be recast as billiards problems. A billiard trajectory is called a periodic orbit if the number of distinct collisions in the trajectory is finite. We show that periodic orbits on such billiard tables cannot have an odd number of distinct collisions. We classify all possible equivalence classes of periodic orbits on square and rectangular tables. We also present a connection between the number of different equivalence classes and Euler's totient function, which for any positive integer N, counts how many positive integers smaller than N share no common divisor with N other than 1. We explore how to construct periodic orbits with a prescribed (even) number of distinct collisions and investigate properties of inadmissible (singular) trajectories, which are trajectories that eventually terminate at a vertex (a table corner).

2020 *Mathematics subject classification*: primary 37C83; secondary 37E15, 37C55.

Keywords and phrases: periodic orbit, mathematical billiard, Euler's totient function.

1. Introduction

Mathematical billiards shares many similarities to the game of billiards in reality. Fundamentally, both comprise a billiard ball moving in a straight line until it strikes a side of the table. Mathematical billiards ignores friction, takes the billiard ball as

[1]Department of Mathematics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand; e-mail: hche435@aucklanduni.ac.nz, h.m.osinga@auckland.ac.nz

a point mass and assumes purely elastic collisions. As a consequence, the billiard trajectory will satisfy the law of reflection (*the angle of incidence equals the angle of reflection*) when it collides with a boundary. Therefore, any billiard trajectory is uniquely determined by its initial position and direction of motion. A real billiard table is almost always rectangular, but mathematical billiard tables can have arbitrary shapes and dimensions. First posed as Alhazen's problem in optics by Ptolemy in 150 AD [4], mathematical billiards is much more than a fun and interesting game; it turns out that the mere shape of the table distinguishes mathematical billiards into three different classes—elliptic, hyperbolic and parabolic—which are well known as classes in different fields of mathematics, not the least in dynamical systems theory [35, 49]. Mathematical billiards has extensively been studied from the perspective of ergodic theory in dynamical systems, as well as algebraic geometry (moduli spaces) and Teichmüller theory; for example, see the textbooks by Tabachnikov [52], Chernov and Markarian [15] or Rozikov [45]. Nevertheless, a myriad of unsolved open problems in mathematical billiards remain [8, 19, 27, 32]. Indeed, there are many fundamental theory questions in mathematics that can be recast as a billiard problem [12, 50], as well as a plethora of applications, including but not limited to the computation of π [25], mechanics [1], quantum computing [24], pouring problems [43], Benford's law [52], diffusion in the Lorentz gas [20] and the Riemann hypothesis [11].

For the class of planar, polygonal tables, that is, a flat shape bounded by a piecewise-linear closed curve, any billiard trajectory exhibits one of only three possible behaviours:

(1) it is a *singular orbit*—after a finite number of collisions, the billiard trajectory terminates at one of the vertices of the table;
(2) it is a *periodic orbit*—after a finite number of collisions, the billiard trajectory retraces itself;
(3) it is a *nonperiodic orbit*—the billiard trajectory continues indefinitely, without ever retracing itself.

Obviously, singular orbits only exist on tables with vertices and they are often ignored or classified as nonperiodic [10]. If a billiard trajectory is periodic, its period is given by the total number of unique collisions. A natural question to ask is whether periodic orbits always exist or whether this depends on the shape of the billiard table. Furthermore, what periods are possible? In fact, the very question whether periodic orbits exist on any polygon is still an open problem! It is listed as Problem 3 in Katok's five most resistant problems in dynamics [34]. Even the case of a triangular billiard table is considered impenetrable, although the triangular billiards conjecture, which states that every triangular table has a periodic billiard path, is widely believed to be true [48]. In particular, a so-called *Fagnano orbit* can readily be constructed in any acute triangle as the period-three orbit formed by the inscribed triangle of minimal perimeter (the orthic triangle); the collision points of such a Fagnano orbit are given by the points at which the altitudes of the triangular billiard intersect the opposite sides [16, 51, 54].

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

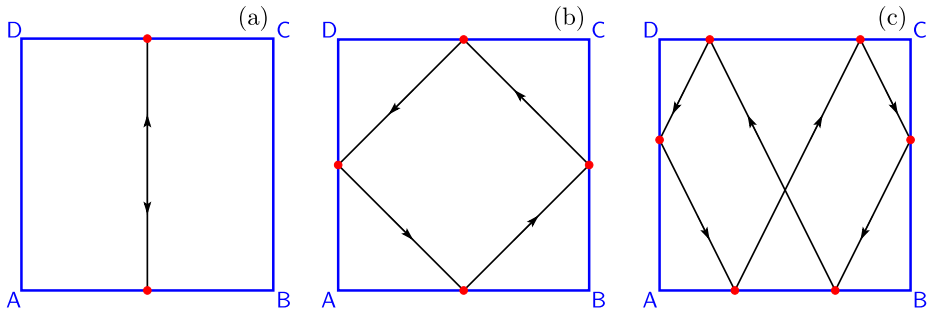

FIGURE 1. Examples of periodic orbits on the square billiard table; the orbits in panels (a), (b) and (c) have periods two, four and six, respectively.

The question about which periods are possible has been treated for specific polygonal billiard tables. In particular, Baxter and Umble [5] classified all possible periodic orbits for an equilateral triangle. In this paper, we consider square or rectangular billiard tables and we similarly give a complete classification of which periods and types of periodic orbits are possible for a square or rectangle. The square is perhaps the simplest piecewise-smooth billiard table, which means that it is introduced early on in billiard textbooks, typically alongside the smooth, circular billiard; for example, see [15, 45, 52]. As a consequence, results for the square billiard tend to be skimmed over, despite its behaviour being distinctly different from circular and other smooth, convex tables. Existence of periodic orbits on the square and rectangular billiard is easily verified by explicit examples. Figure 1 shows three examples of periodic orbits on the square. Here, the square table is represented by the (blue) boundary with vertices A, B, C and D, the periodic orbits are indicated as directed (black) lines that collide with the boundary at points identified by thick (red) dots. By counting these points of collision, we see that the periods are two, four and six in the respective panels (a), (b) and (c). Note that the period-two orbit in panel (a) is readily produced on a rectangle as well; for the period-four and -six orbits, this is less obvious.

Interestingly, none of the textbooks show that the square and rectangular billiard tables do not admit periodic orbits with odd periods. This can be proven using cutting sequences, as done by Davis [17], but we provide a different proof that harmonizes well with the periodicity classification presented in this paper. Furthermore, we also discuss how to construct periodic orbits with prespecified periods and connect our classification of periodic orbits with Euler's totient function from number theory [38].

From a practical point of view, our findings may be of interest for the study of so-called integer periodic billiard orbits for a special class of rectangular billiards, such that all collisions occur at integer lattice points [22]; rectangular lattice billiards are particularly relevant for discrete tomography [28]. Furthermore, there is a renewed interest in the phenomenon known as *quantum scars*, when a classical periodic billiard trajectory appears as the approximation of a quantum mechanical wave function for an otherwise classically chaotic quantum system [6]. The techniques of selecting a

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

specific wave function to observe a particular quantum scar is an ongoing experimental challenge, even already for the classical Bunimovich stadium [26]. The direct relation to quantum mechanical wave functions has fuelled interest in a diverse range of particle transport problems related to square or polygonal billiards [31]; specific examples are particle transport in infinitely long channels with a right-angled corner [41] and experimental patterns generated in a square-shaped microcavity laser [14].

This paper is organized as follows. In the next section, we define the different classes of periodic orbits and clarify what we mean when two periodic orbits with the same period are different. Here, we also explain the powerful technique of unfolding [23, 56] that is used in the classical proof of existence of periodic orbits on the square or rectangular billiard table. We present alternative and intuitive proofs for periodicity and nonexistence of odd periodic orbits in Section 3. In Section 3.2, we count and fully classify all possible periodic orbits. We treat singular orbits in Section 3.3, where we show how different types of singular orbits form boundaries between different families of periodic orbits. In Section 4, we extend the results for the square to the rectangular billiard table. We conclude in Section 5 with a discussion of future work.

2. Setting and definitions

The three periodic orbits shown in Figure 1 for the square billiard table are all different, because their periods are not the same. However, the period-two orbit in Figure 1(a), for example, can be shifted to the right or left without changing its period; similarly, the period-six orbit in Figure 1(c) can be flipped upside down without changing the period. In this section, we define the *family* or *class* of periodic orbits that we consider to be equivalent, and we will then proceed to count the number of different classes of periodic orbits.

We begin by defining a coordinate system that identifies a quadrilateral billiard table, denoted ABCD, with four (ordered) vertices A, B, C and D. Note that the billiard table can (uniformly) be scaled without affecting the number of distinct collisions for a periodic orbit; this means that we may assume that side AB has unit length. Moreover, we consider this side as the base and position the table such that AB is equal to the unit interval $\mathsf{AB} := \{(x, 0) \in \mathbb{R}^2 \mid 0 \leq x \leq 1\}$ on the x-axis, and the vertex A lies at the origin. Recall that any billiard trajectory is uniquely determined by an initial position and associated direction of motion. We assume that the initial position is a point of collision with the side AB; if necessary, we rotate the table and relabel the vertices so that such a point exists. Then, we can identify any billiard trajectory by the initial point of collision on the side AB at distance $P_0 \in [0, 1)$ from A, and an initial angle $\alpha_0 \in (0, \pi/2]$ between the corresponding outgoing (or equivalently incoming) line and the side AB; we will say that the pair $\langle P_0, \alpha_0 \rangle$ generates the billiard trajectory. Note that we restrict the angle α_0 to be at most $\pi/2$ radians, which means that we do not specify the direction of motion. Since we are primarily interested in periodic orbits, and how many different ones there are, we do not care whether the sequence of collision points is given in reversed order. We will slightly abuse notation and refer to P_0 as both the initial point on AB and its coordinate on the x-axis.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

We only consider square or rectangular billiard tables; hence, the adjacent sides are perpendicular, which has an important consequence.

PROPOSITION 2.1. *Consider a billiard trajectory on the square or rectangular billiard* ABCD *generated by the pair* $\langle P_0, \alpha_0 \rangle$ *with* $P_0 \in [0, 1)$ *and* $\alpha_0 \in (0, \pi/2]$. *Then, any collisions with sides* AB *and* CD *will be at angle* α_0 *and any collisions with sides* BC *and* DA *will be at angle* $\pi/2 - \alpha_0$.

PROOF. Since $\alpha_0 \in (0, \pi/2]$, the outgoing line from P_0 on the side AB will either end on the adjacent side BC or on the opposite side CD. The sides AB and CD are parallel, because ABCD is either a square or a rectangle. Hence, if the outgoing line from P_0 ends on CD, it does so at the same angle α_0; see Figures 1(a) and 1(c) for examples. If the outgoing line from P_0 ends on the adjacent side BC, then it does so at an angle α_1 such that $\alpha_0 + \pi/2 + \alpha_1 = \pi$, because the outgoing line forms a right triangle with the sides AB and BC; see Figures 1(b) and 1(c) for examples. Therefore, we have $\alpha_1 = \pi/2 - \alpha_0$ as claimed. The end point of the outgoing line from P_0 is another collision point P_1, unless it is a vertex, in which case, the billiard trajectory terminates and the result holds. By rotating the table and relabelling the vertices, we can consider the billiard trajectory as being generated by the pair $\langle P_1, \alpha_1 \rangle$, with $\alpha_1 = \alpha_0$ or $\alpha_1 = \pi/2 - \alpha_0$. Using the same arguments, we find that a collision with the (original) side DA also occurs at angle $\pi/2 - \alpha_0$. It follows that all collisions with sides AB and CD occur at angle α_0, and all collisions with sides BC and DA occur at angle $\pi/2 - \alpha_0$. Note the special case when $\alpha_0 = \pi/2$, which leads to a collision with the opposite side CD at a point P_1 that lies directly above P_0. Hence, the next collision will again be with side AB, at the same point P_0, and the billiard trajectory is a period-two orbit that never collides with the sides BC and DA; see the example in Figure 1(a). □

We use Proposition 2.1 to classify different billiard trajectories with the same period: we distinguish them by the numbers of different collision points in the cycle that correspond to the two different angles at which these collisions occur. More precisely, we have the following definition of equivalence.

DEFINITION 2.2 (Equivalence class $C_K(p)$ of period-K orbits). A period-K orbit for the square (or rectangular) billiard ABCD belongs to the equivalence class $C_K(p)$, for some $0 < p < K$, if a total of p of its K different collision points lie on the (parallel) sides AB or CD. In the special case $K = 2$, we define the equivalence classes $C_2(2)$ and $C_2(0)$ of period-two orbits that exclusively collide with the pair of parallel sides AB and CD or BC and DA, respectively.

REMARK 2.3. The class of period-two orbits is special. There are no other periodic orbits that only collide with one pair of parallel sides. Recall that the initial position of any billiard trajectory is assumed to be a point of collision with the side AB. Hence, technically, the class $C_2(0)$ is not even included in our setting; in what follows, we will treat the period-two orbits separately. For all other periodic orbits, the total number p of collisions with the sides AB and CD and the total number q of collisions with the

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

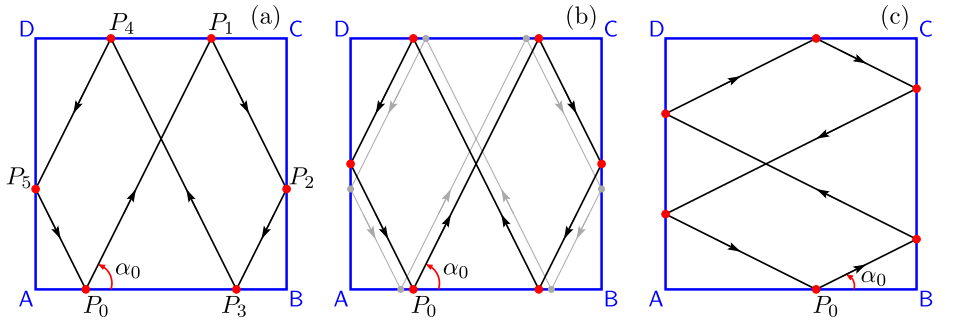

FIGURE 2. Examples of period-six orbits on the square billiard, generated by (a) the initial pair $\langle P_0, \alpha_0\rangle = \langle 0.2, \tan^{-1}(2)\rangle$, (b) the shifted initial pair $\langle P_0, \alpha_0\rangle = \langle 0.25, \tan^{-1}(2)\rangle$ that is from the same equivalence class, and (c) the pair $\langle P_0, \alpha_0\rangle = \langle 0.6, \pi/2 - \tan^{-1}(2)\rangle$ obtained by rotation, which we consider part of a different family of period-six orbits.

sides BC and DA satisfy $p \geq 1$ and $q \geq 1$; we write $p, q \in \mathbb{N}$ with the understanding that $\mathbb{N}$ represents the set of *positive* natural numbers.

The three different periodic orbits shown in Figure 1 with periods two, four and six can be distinguished more precisely as periodic orbits from the three different equivalence classes $C_2(2)$, $C_4(2)$ and $C_6(4)$, respectively.

Note that any period-K orbit from the class $C_K(p)$ has $q = K - p$ different collision points on the (parallel) sides BC or DA, which we may emphasize by saying the period-K orbit is of type (p, q). We illustrate this by comparing the period-six orbits in Figure 2 with the possibly equivalent period-six orbit from Figure 1(c). We identify collision points by their coordinates on the boundary of the unit square, that is, we assume vertices A and B have coordinates $(0, 0)$ and $(1, 0)$, and by definition, vertices C and D then have coordinates $(1, 1)$ and $(0, 1)$, respectively.

EXAMPLE 2.4. Figure 2(a) shows the period-six orbit of Figure 1(c), but reflected about the line $\{y = 0.5\}$. More precisely, this period-six orbit is generated by the pair $\langle P_0, \alpha_0\rangle = \langle 0.2, \tan^{-1}(2)\rangle$, which means that it starts at the point $P_0 = (0.2, 0)$ on the side AB along the outgoing line with slope 2. The sequence of successive collision points is then given by the points $P_1 = (0.7, 1)$, $P_2 = (1, 0.4)$, $P_3 = (0.8, 0)$, $P_4 = (0.3, 1)$, $P_5 = (0, 0.4)$, after which the cycle repeats with P_0. Note that all lines are parallel to either the outgoing line at P_0, with slope 2, or the incoming line at P_0, with slope -2. Hence, the angles at P_0 and P_3 on the horizontal side AB are both α_0, as are the angles at P_1 and P_4 on the other horizontal side CD; however, with respect to the vertical sides BC and DA, these lines only have slopes $\pm 1/2$ so that the angles at P_2 and P_5 are $\tan^{-1}(1/2) = \pi/2 - \alpha_0$. Therefore, this period-six orbit is of type $(4, 2)$ and it is equivalent to the period-six orbit from Figure 1(c). If we choose the reversed direction of motion, starting along the line with slope -2, we encounter the reversed sequence of collisions P_5, P_4, P_3, P_2, P_1 and P_0, leading again to an equivalent periodic orbit of type $(4, 2)$.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

EXAMPLE 2.5. Figure 2(b) shows another period-six orbit of type $(4, 2)$. It is similar to the one shown in panel (a), but the point P_0 is shifted further along the side AB, to the point $(0.25, 0)$, while maintaining the same angle α_0. A further shift of P_0 to $(0.3, 0)$ leads to the period-six orbit from Figure 1(c). Indeed, these shifted periodic orbits are all part of a family of period-six orbits in $C_6(4)$ that is generated by a pair $\langle P_0, \alpha_0 \rangle$, with P_0 varying over one or more sub-intervals in $[0, 1]$.

EXAMPLE 2.6. Figure 2(c) shows the same period-six orbit from panel (a), but rotated anti-clockwise by a quarter turn; a mere relabelling of the vertices leads to a periodic orbit generated by the pair $\langle P_5, \tan^{-1}(1/2) \rangle = \langle 0.6, \tan^{-1}(1/2) \rangle$, which is an example from a different equivalence class, namely, from $C_6(2)$.

In the setting for this paper, the periodic orbit of type $(2, 4)$ shown in Figure 2(c) is not from the same equivalence class as the periodic orbits of type $(4, 2)$ shown in panels (a) and (b). We make a distinction between these two equivalence classes, because it would be natural to do so for rectangular billiard tables.

2.1. Unfolding of billiard trajectories Following a billiard trajectory on a table ABCD can be difficult if there are many collisions, because of a myriad of intersecting lines. The technique of *unfolding* transforms the billiard trajectory: rather than reflecting collisions with the sides, the entire table reflects, so that the billiard trajectory remains a straight line. For so-called *rational*, planar, polygonal billiard tables, which have polygonal angles that are rational multiples of π, this approach relates mathematical billiards to the theory of geodesic flow on Riemann surfaces [9, 23, 56]; in special cases of rational billiard tables, including the square and rectangle, this unfolding leads to a tiling of the plane and is amenable to illustration.

EXAMPLE 2.7. We show an example in Figure 3 with an unfolding of the period-six orbit from Figure 2(c). Recall that this periodic orbit is generated by the pair $\langle P_0, \alpha_0 \rangle = \langle 0.6, \tan^{-1}(1/2) \rangle$ on the square ABCD and there are six different collisions, which are the points $P_0 = (0.6, 0)$ on the side AB, followed by $P_1 = (1, 0.2)$, $P_2 = (0, 0.7)$, $P_3 = (0.6, 1)$, $P_4 = (1, 0.8)$ and $P_5 = (0, 0.3)$. Here, P_1 and P_4 lie on the side BC, and P_2 and P_5 lie on the side DA, while P_3 lies on CD. In the unfolding, the trajectory from P_0 to P_1 continues in a straight line with slope $1/2$ beyond the side BC, on the (horizontally) reflected table BADC; the next collision occurs on the opposite side, which is AD as required, at the same location P_2 as would have been reached by following the outgoing line with slope $-1/2$ from P_1 in the original table ABCD. At P_2, the table is again reflected horizontally and the trajectory continues in the original orientation until the collision at P_3 on the side CD. Next, the trajectory continues with the same slope beyond CD, on the vertically reflected table DCBA; the subsequent collision points P_4 and P_5, together with two associated vertical reflections, are oriented up-side-down with respect to the original table ABCD. Indeed, the trajectory from P_4 to P_5 continues at the same angle $\tan^{-1}(1/2)$ with the horizontal, because in the unfolding, it lies on the table CDAB. The horizontal reflection used to continue

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

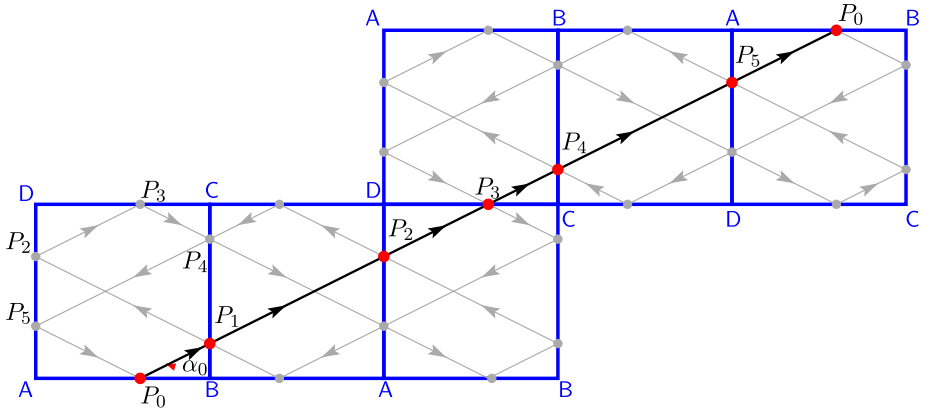

FIGURE 3. The unfolded period-six orbit from Figure 2(c) on the square ABCD. The periodic orbit has type $(2,4)$, which means that the unfolding requires two reflections about a horizontal side and four reflections about a vertical side before the trajectory repeats on a translated copy of tables with the same orientations.

past P_5 flips the table back to the orientation DCBA that was used to pass from P_3 to P_4. The next collision with the side AB at the top right in Figure 3 corresponds to the point P_0 on AB; the table obtained after a sixth reflection about the side AB will be a translated copy of the original table ABCD on the bottom left, and the trajectory will repeat. Hence, the line segment in Figure 3 can be continued past P_0 in both directions to give a straight line in $\mathbb{R}^2$ that has slope $1/2$ and passes through the point P_0 on the x-axis; the entire line is the unfolded trajectory in the plane, which in this example represents a period-six orbit of type $(2,4)$.

REMARK 2.8. The technique of unfolding for the rectangular billiard is done in the exact same way with horizontal and vertical reflections. Since the direction of motion at a point of collision only depends on the angle it makes with that side of the table, trajectories on a rectangular billiard also correspond to straight lines in $\mathbb{R}^2$ determined by the corresponding generator $\langle P_0, \alpha_0 \rangle$.

3. Properties of the square billiard

Notice in Figure 3 that the orientation of the table changes each time a collision occurs. More precisely, there are four unique orientations, namely, the original table ABCD, the horizontally reflected version BADC, the vertically reflected version DCBA, and the table CDAB that is both horizontally and vertically reflected with respect to the original table. We can identify these four orientations in terms of the integer coordinates of their vertices in $\mathbb{R}$.

DEFINITION 3.1. Consider a square tile in $\mathbb{R}^2$ with bottom-left vertex $(i,j) \in \mathbb{Z} \times \mathbb{Z}$. Then this tile corresponds to only one of the following four orientations:

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

- if both i and j are even, then the tile corresponds to ABCD, which has positive horizontal and positive vertical orientation;
- if i is odd and j is even, then the tile corresponds to BADC, which has negative horizontal and positive vertical orientation;
- if both i and j are odd, then the tile corresponds to CDAB, which has negative horizontal and negative vertical orientation;
- if i is even and j is odd, then the tile corresponds to DCBA, which has positive horizontal and negative vertical orientation.

By combining all four orientations together, we can also view $\mathbb{R}^2$ as being tiled with squares that have sides ADA and ABA, twice the length of those for the original table ABCD. The tiling with such larger squares is done by mere translations, rather than reflections. Consequently, this larger square is a representation of the (flat) torus $\mathbb{T}^2$, as illustrated in Figure 4.

EXAMPLE 3.2. Figure 4(a) shows again the unfolding of the period-six orbit from Figure 2(c), but as this billiard trajectory disappears off the edge of the square on the right-hand side ADA, it reappears on the left-hand side ADA and continues with the same slope; similarly, when the top side ABA is reached, the trajectory reappears on the bottom side ABA. Hence, the period-six orbit is now represented by a set of three parallel straight-line segments, rather than a single straight line. These line segments form a single curve if we identify the left- and right-hand sides ADA by folding the square into a cylinder; see Figure 4(b). If we now also identify the top and bottom sides ABA, as done in Figure 4(c), then the period-six orbit is, in fact, given by a closed curve on $\mathbb{T}^2$.

The alternative representation in Figure 4 is the basis for the classical proof of existence and classification of periodic orbits on the square billiard, which uses the notion of *geodesic* on $\mathbb{T}^2$, that is, a length-minimizing curve. In the usual Euclidean geometry, geodesics are straight lines, and $\mathbb{T}^2$ inherits this geometry by way of the construction illustrated in Figure 4. The theory of geodesics tell us that all straight lines with rational slopes correspond to closed geodesics on $\mathbb{T}^2$; for example, see [42]. Therefore, if a trajectory on the square billiard is periodic, then the slope must be rational.

REMARK 3.3. The theory of geodesics also states that all straight lines with irrational slopes correspond to geodesics that densely fill the torus. This dichotomy is known as *Veech dichotomy* in the literature; it implies that the square billiard is *ergodically optimal* [55]. We refer again to [42] for details.

The classical proof has the advantage that it extends readily to higher dimensions by using analogous arguments [52]. However, this approach ignores the possibility that a billiard trajectory with rational slope terminates at one of the four vertices; indeed, the four corners of the billiard table have no special role on $\mathbb{T}^2$ and any geodesic can (uniquely) pass through such points. Hence, one could opt to ignore the possibility of

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

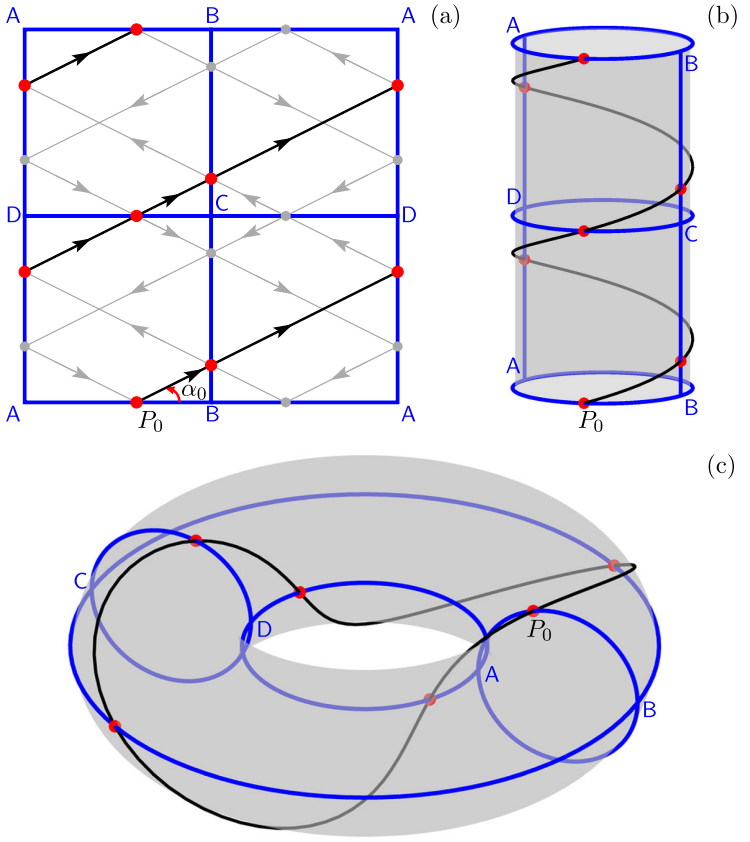

FIGURE 4. Equivalent representation of the period-six orbit from Figure 2(c) on the torus. Panel (a) illustrates the trajectory on the large square table comprising all four orientations of the square ABCD. The left and right sides ADA are identified to form a cylinder in panel (b), and the top and bottom sides ABA are subsequently identified to form the torus in panel (c), on which the period-six orbit forms a closed curve.

a singular orbit, because it is possible to extend this billiard trajectory after hitting the corner in a well-defined manner that is continuous with respect to initial conditions. In this paper, however, we analyse the *trichotomy* that includes the possibility of singular orbits for a square billiard. In this setting, the condition that the slope be rational is not sufficient to guarantee periodicity. We use the technique of unfolding in $\mathbb{R}^2$ as an alternative and more intuitive way to prove important properties of the square billiard, such as the nonexistence of periodic orbits with odd periods [17] and what the period of a billiard trajectory will be for a given rational slope. We will also count and classify all different equivalence classes of periodic orbits that have a given period associated with a particular rational slope. We find that precisely the singular orbits act as separatrices of these different equivalence classes.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

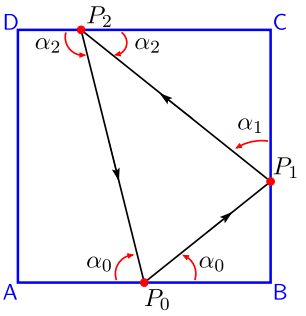

FIGURE 5. Hypothetical period-three orbit on the square ABCD shown as a triangle composed of the three points of collision P_0, P_1 and P_2 at which the billiard trajectory makes angles α_0, α_1 and α_2 with the table, respectively. The incoming and outgoing angles at these points are intended to be equal, and labelled so accordingly.

3.1. Periods of periodic orbits for the square billiard are even All the examples of periodic orbits for the square billiard given so far have even periods. This is not a coincidence. In particular, is it not possible to have a periodic orbit with period three.

PROPOSITION 3.4. *The square billiard does not admit a period-three orbit.*

PROOF. Suppose for the sake of contradiction that a period-three orbit exists. To focus the mind, Figure 5 illustrates this hypothetical billiard trajectory on the square ABCD. The three different points of collision are denoted P_0, P_1 and P_2, and their associated angles are α_0, α_1 and α_2, respectively; here, angles with the same label are supposed to be equal, even though the image may suggest otherwise. Note that any period-three orbit must collide with three different sides of the square, because it is impossible for a billiard trajectory to incur two consecutive collisions on the same side. Since we can always rotate the table and relabel the vertices, the illustration in Figure 5 is representative for any period-three orbit. In other words, without loss of generality, we may assume that P_0 lies on the side AB, and the other two successive points of collision P_1 and P_2 lie on the sides BC and CD, respectively. Recall from Section 2 that the billiard trajectory is uniquely generated by the pair $\langle P_0, \alpha_0 \rangle$, and Proposition 2.1 implies that we must have $\alpha_1 = \pi/2 - \alpha_0$ and $\alpha_2 = \alpha_0$. Furthermore, the sum of the angles in the quadrilateral formed by P_0, P_2, and the two vertices D and A of the square table must be 2π, which implies that $\alpha_0 + \alpha_2 = \pi$. Therefore, if a period-three orbit exists, then

$$\left.\begin{aligned} \alpha_0 + \alpha_2 &= \pi \\ \alpha_2 &= \alpha_0 \end{aligned}\right\} \Longrightarrow 2\alpha_0 = \pi \Longleftrightarrow \alpha_0 = \frac{\pi}{2}.$$

However, the pair $\langle P_0, \pi/2 \rangle$ generates a period-two orbit for all $P_0 \in (0, 1)$, because the billiard trajectory will bounce back and forth between the two points P_0 and P_2; which is a contradiction. □

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

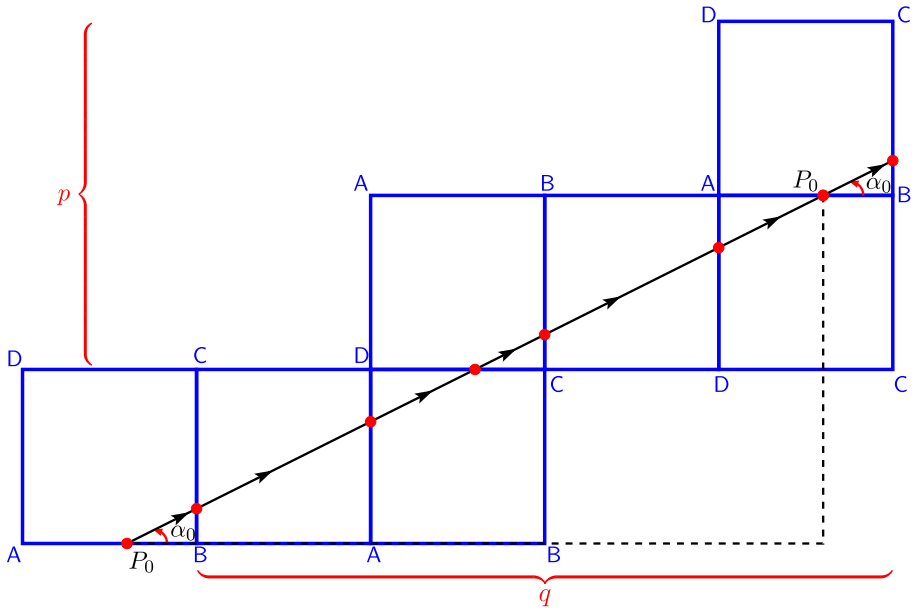

FIGURE 6. Unfolding on $\mathbb{R}^2$ of a periodic orbit for the square ABCD with period $p + q$ composed of p vertical and q horizontal reflections before returning to the initial position P_0.

We cannot easily extend this geometric proof to other odd periods as the number of different cases grows very quickly and self-intersections cause difficulty to make the same arguments. Therefore, we use the technique of unfolding, explained in Section 2.1, to prove the following general result.

THEOREM 3.5. *Any periodic orbit for the square billiard has even period and is generated by a pair* $\langle P_0, \alpha_0 \rangle$*, with* $P_0 \in [0, 1)$*, and either* $\alpha_0 = \pi/2$ *or* $\alpha_0 \in (0, \pi/2)$ *such that* $\tan(\alpha_0)$ *is rational.*

PROOF. As was already mentioned in the proof of Proposition 2.1, the case $\alpha_0 = \pi/2$ generates a period-two orbit; hence, the period for this special case is even. This case is special, because the slope $\tan(\alpha_0)$ of the outgoing line is not defined for $\alpha_0 = \pi/2$. We now consider a periodic orbit for the square ABCD generated by a pair $\langle P_0, \alpha_0 \rangle$ with $P_0 \in [0, 1)$ and $\alpha \in (0, \pi/2)$. Note that $P_0 \neq 0$, because a periodic billiard trajectory that starts at a vertex must also return to this vertex, which would make it a singular orbit. The unfolding on $\mathbb{R}^2$ of such a periodic orbit is a straight line with slope $\tan(\alpha_0)$ that passes through the point at distance P_0 from the vertex A at the origin, and such that it intersects the tiling of $\mathbb{R}^2$ along other (although not necessarily all) sides AB at distances P_0 from the shifted vertex A in translated copies of the original table ABCD.

An example is shown in Figure 6, where again the period-six orbit from Figure 2(c) was used. The slope $\tan(\alpha_0)$ of the unfolded trajectory is determined by the difference between the x- and y-coordinates of the initial point P_0 and its translated copy; this

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

is suggestively indicated by the dashed triangle in Figure 6. Suppose that the periodic orbit is of type (p, q) for some $p, q \in \mathbb{N}$, which means that the unfolded trajectory passes through a translated copy of P_0 (with the correct orientation) for the *first* time after p vertical and q horizontal reflections. Since the distance between P_0 and the vertex A is preserved in the translated copy, the line through $(0, 0)$ and (q, p) in $\mathbb{R}^2$ is parallel to the unfolded trajectory. Hence, the slope of the unfolded trajectory is $\tan(\alpha_0) = p/q$, which is rational, as claimed.

Furthermore, Definition 3.1 states that a translated copy of the billiard table with orientation ABCD is only obtained when the (integer) coordinates of the shifted vertex A are both even, that is, p and q must both be even. The period of the periodic orbit is given by the total number of distinct collisions; we claim that this number is $p + q$, and since p and q are even, the period of the periodic orbit is even. It remains to prove that the total number, say K, of distinct collisions is $p + q$. Recall that the technique of unfolding transforms the billiard trajectory into a straight line by reflecting the table at points of collision instead of changing the direction of motion. Hence, K is at most equal to the number of reflections, that is, $K \leq p + q$. Now, suppose $K < p + q$, which means that the billiard trajectory is repeated after K collisions. However, p and q are, by definition, the number of horizontal and vertical reflections needed to pass through a translated copy of P_0 with the correct orientation for the *first* time. Hence, $P_\ell \neq P_0$ for all $0 < \ell < p + q$; therefore, $K = p + q$ and the period is even. □

In what follows, we characterize all periodic orbits of period K, for any positive integer $K \in 2\mathbb{N}$, in terms of the conditions we must impose on $P_0 \in [0, 1)$ and $\alpha \in (0, \pi/2]$ such that the pair $\langle P_0, \alpha_0 \rangle$ gives rise to such a period-K orbit for the square billiard ABCD. In particular, we confirm that periodic orbits exist for all even periods, and explain how many different equivalence classes there are.

3.2. Construction of periodic orbits for the square billiard If the trajectory generated by a pair $\langle P_0, \alpha_0 \rangle$ is periodic, then the unfolded trajectory will intersect shifted copies of the side AB infinitely many times at distances P_0. The (rational) slope $\tan(\alpha_0)$ of this trajectory can be determined using any such shifted copy: if a periodic orbit is obtained after p horizontal and q vertical reflections, then the same periodic orbit is obtained after dp horizontal and dq vertical reflections, for any $d \in \mathbb{N}$, because $\tan(\alpha_0) = p/q = dp/dq$. For the same reason, if $d \in \{3, \ldots, \min(p, q)\}$ divides both p and q, then there exists an intersection at distance P_0 on a shifted copy of the side AB that lies closer to the original AB. Of course, $d = 2$ divides p and q, because they are both even, but the corresponding point $(P_0 + p/2,\ q/2)$ does not necessarily lie on the side AB, or if it does, this side may have the reflected orientation BA instead. Inspired by Baxter and Umble [5] who give a similar definition for the equilateral triangle, we formalize the notion of *least* period of a periodic unfolded trajectory.

DEFINITION 3.6. Given an unfolded trajectory on $\mathbb{R}^2$ that is periodic after p horizontal and q vertical reflections with $p, q \in \mathbb{N}$ and $\gcd(p, q) = 2$, then its *least* period is $p + q$.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

To generate a periodic orbit with least period $K \in 2\mathbb{N}$, we first construct the corresponding unfolded trajectory in $\mathbb{R}^2$. This unfolded trajectory is given by a line with slope p/q, where $p \in \mathbb{N}$ and $q \in \mathbb{N}$ are chosen such that $\gcd(p, q) = 2$ and $p + q = K$. In fact, since both p and q are even, we can look for $m = p/2$ and $n = q/2$ such that n and m are co-prime and $m + n = K/2$; indeed, $p/q = m/n$ and $\gcd(p, q) = 2\ \gcd(m, n) = 2$.

EXAMPLE 3.7. To generate a periodic orbit with least period $K = 4$, for example, as shown in Figure 1(b), we must find $m, n \in \mathbb{N}$ with $\gcd(n, m) = 1$ such that $m + n = K/2 = 2$. The only such candidate is $n = m = 1$. Therefore, any period-four orbit is generated by a pair $\langle P_0, \alpha_0 \rangle$ with $\alpha_0 = \tan^{-1}(1) = \pi/4$. Here, the choice for $P_0 \in [0, 1)$ is arbitrary, except that we must avoid generating a singular orbit; for the period-four orbit, it is perhaps not hard to see that the only restriction is $P_0 \neq 0$, but see Section 3.3 for details.

Note that for any other even period K, two easy choices are $m = 1$ and $n = K/2 - 1$, or vice versa, which both satisfy Definition 3.6. Hence, as long as $m \neq n$, this trivial decomposition already generates two different periodic orbits with the same period: one periodic orbit is of type $(2, 2n)$ and the other of type $(2m, 2)$, which means, according to Definition 2.2, that these periodic orbits are from the two different equivalence classes $C_K(2)$ and $C_K(2m) = C_K(K - 2)$, respectively.

For sufficiently large K, we expect there to be other pairs of co-prime numbers that sum to $K/2$, that is, we expect there to be more equivalence classes and different types of periodic orbits with the same period $K \in 2\mathbb{N}$. To determine how many possible decompositions there are, we use Euler's totient function from number theory [38]. For any $N \in \mathbb{N}$, Euler's totient function $\varphi(N)$ counts the number of natural numbers $m \in \{1, \ldots, N\}$ such that $\gcd(m, N) = 1$. The positive integers m that satisfy this property are referred to as *totatives* of N.

PROPOSITION 3.8. *The total number of ordered pairs (m, n) with $m, n \in \mathbb{N}$ and $m + n = N$, such that $\gcd(m, n) = 1$ is equal to Euler's totient function $\varphi(N)$.*

PROOF. Consider $m, n \in \mathbb{N}$ with $\gcd(m, n) = 1$ and define $N = m + n$. By definition, $\gcd(m, n) = \gcd(m, m + n) = \gcd(m, N)$. However, $\gcd(m, n) = 1$, so $\gcd(m, N) = 1$, which means that m is a totative of N. Therefore, each ordered pair (m, n) is determined by whether m is a totative of N, and the total number of possibilities is $\varphi(N)$. □

The total of $\varphi(N)$ different pairs (m, n) lead to $\varphi(N)$ different ratios m/n, so that we may conclude the following.

COROLLARY 3.9. *For $K = 2N$, with $N \in \mathbb{N}$, there are $\varphi(N)$ different types of period-K orbits for the square billiard, generated by $\varphi(N)$ different angles in the interval $(0, \pi/2]$, which represent a total of $\varphi(N)$ different equivalence classes.*

REMARK 3.10. Note that Corollary 3.9 includes the special case of a period-two orbit: for $K = 2$, we have $N = 1$ and $\varphi(N) = 1$. Indeed, there is only one type of period-two

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

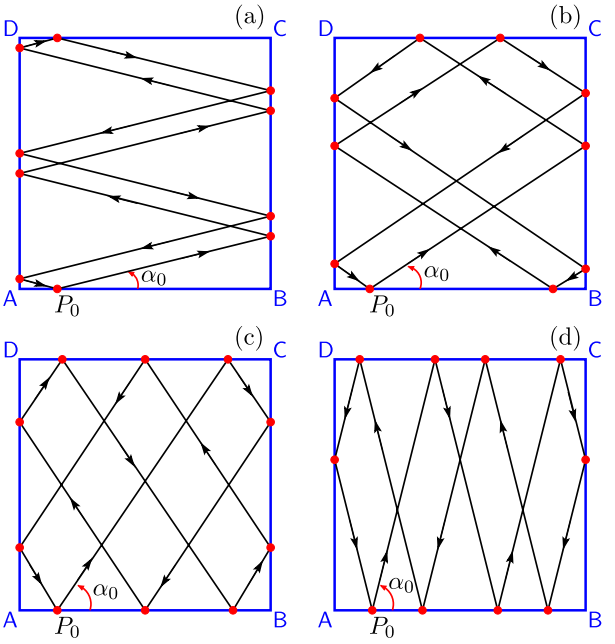

FIGURE 7. Period-ten orbits generated from $P_0 = 0.15$ for the square ABCD from each of the four equivalence classes. The periodic orbits in panels (a)–(d) are of type $(2, 8)$, $(4, 6)$, $(6, 4)$ and $(8, 2)$, respectively.

orbit for the square billiard, namely, the one generated by the angle $\alpha_0 = \pi/2$. If we also were to consider the family of period-two orbits with 0 collision points on the parallel sides AB and CD, then there would actually be two types of period-two orbits, leading to the two different equivalence classes $C_2(2)$ and $C_2(0)$.

EXAMPLE 3.11. There exist $\varphi(5) = 4$ different types of period-ten orbits. Note that 5 is prime, so the totatives of 5 are all positive integers 1, 2, 3, 4 less than 5. Therefore, there are four possible ordered pairs, namely, $(1, 4)$, $(2, 3)$, $(3, 2)$ and $(4, 1)$, leading to four different families of period-ten orbits from the corresponding equivalence classes $C_{10}(2)$, $C_{10}(4)$, $C_{10}(6)$ and $C_{10}(8)$, respectively. Figure 7 illustrates these four different types of period-ten orbits generated by the pairs $\langle P_0, \alpha_0 \rangle$ with $P_0 = 0.15$ fixed and $\tan(\alpha_0) = 1/4, 2/3, 3/2$ and $4/1$, in panels (a)–(d), respectively.

EXAMPLE 3.12. Corollary 3.9 implies that there cannot exist a period-eight orbit of type $(4, 4)$, that is, a periodic orbit that unfolds into a trajectory which repeats after four vertical and four horizontal reflections. Indeed, the corresponding slope would have to be 1 and, thus, $\alpha_0 = \tan^{-1}(1) = \pi/4$, which generates a period-four orbit. Hence, such a period-eight orbit would just be a double copy of the period-four orbit; while its period may be eight, the *least* period of this periodic orbit is four.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

REMARK 3.13. There is no efficient algorithm to find the totatives of a given integer $N \in \mathbb{N}$, because it is equivalent to prime factorization. Indeed, Euler's totient function is explicitly given by the classical Euler's product formula,

$$\varphi(N) = N \prod_{p|N} \left(1 - \frac{1}{p}\right),$$

which computes a product over all distinct prime factors of N. An alternative formula,

$$\varphi(N) = \sum_{k=1}^{N} \gcd(k, N) \cos\left(2\pi \frac{k}{N}\right),$$

can be derived using the discrete Fourier transform [46]; unfortunately, finding the greatest common divisors $\gcd(k, N)$ for all $k \in \{1, \ldots, N\}$ is also computationally as complex as finding prime factors.

3.3. Singular orbits for the square billiard So far, we have not imposed any conditions on the choice for $P_0 \in [0, 1)$. Indeed, with our definition of equivalence class, a periodic orbit generated by $\langle P_0, \alpha_0 \rangle$ is equivalent to other periodic orbits generated by the same initial angle α_0, but starting from a different initial point P_0. However, we cannot choose just any value for P_0. For example, we have already seen that $P_0 = 0$ leads to a singular rather than a periodic orbit. Perhaps there are other values $P_0 \in [0, 1)$ for which the pair $\langle P_0, \alpha_0 \rangle$ generates a billiard trajectory that terminates at one of the vertices? This question was addressed in the very first published article on rectangular (and, thus, also square) billiards by Lennes [37], who treated singular orbits as periodic orbits by defining collisions with a vertex as reflections back along the line of approach. The discussion in [37] centres on billiard trajectories that start at a vertex and whether, for given angle α_0, such a billiard trajectory may eventually return to this vertex. Lennes also derived the required number of collisions in dependence on α_0. As the treatise in [37] is short and lacks complete proofs, we repeat some of the results here.

PROPOSITION 3.14. *If a billiard trajectory starts at a vertex with an irrational slope, then it never terminates at a vertex on the square billiard.*

PROOF. Suppose for the sake of contradiction that there exists a singular orbit that starts from a vertex with irrational slope. Without loss of generality, we may assume that this billiard trajectory starts at vertex A of the square ABCD; if not, we rotate the table and relabel the vertices. We now consider the corresponding unfolded trajectory in the plane $\mathbb{R}^2$. Since we assumed that the billiard trajectory is singular, the unfolded trajectory must pass through a shifted copy of one of the vertices; by Definition 3.1, these shifted copies lie in $\mathbb{Z} \times \mathbb{Z}$. Therefore, the unfolded trajectory contains a line segment from the original vertex A at $(0, 0)$ to a shifted copy of this same or another vertex at the point, say, $(m, n) \in \mathbb{Z} \times \mathbb{Z}$; here, both $n \neq 0$ and $m \neq 0$, because the slope is neither 0 nor ∞. This means that the unfolded trajectory is the straight line with slope

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

$m/n \in \mathbb{Q}$. Any possible rotation back to the original billiard table either preserves this slope (rotation by 0 or π), or changes it to $n/m \in \mathbb{Q}$ (rotation by $\pm\pi/2$), both of which are rational; which is a contradiction. □

COROLLARY 3.15. *If a billiard trajectory for the square billiard* ABCD *that starts at a point* $P_0 \neq$ A *with a rational slope is a singular orbit, then the trajectory generated from* P_0 *in the opposite direction, using the reflected (negative) slope, will also be singular.*

Indeed, both billiard trajectories unfold to the same line with rational slope in the plane $\mathbb{R}^2$. Hence, every singular orbit is contained in a billiard trajectory that both starts and terminates at a vertex. Such billiard trajectories are called *generalized diagonals* [33].

DEFINITION 3.16. A generalized diagonal is a singular orbit that starts at a vertex. The length of a generalized diagonal is given by the total number of (nonvertex) collision points.

We now return to the question whether values $P_0 \in [0, 1)$ for the pair $\langle P_0, \alpha_0 \rangle$ with $\tan(\alpha_0)$ rational, generate periodic rather than singular orbits. As an example, consider the four different period-ten orbits shown in Figure 7.

EXAMPLE 3.17. Notice that the periodic orbit in Figure 7(a), which starts at $P_0 = 0.15$ with slope $1/4$, has five pairs of collision points that are located close together: two pairs on the side BC, one pair on DA, one pair near vertex A on sides AB and DA, and a fifth pair near vertex D on sides DA and CD. If we shift P_0 towards vertex A, these pairs of collision points will move even closer together, until each pair merges as P_0 reaches A. The resulting billiard trajectory, which is shown in Figure 8(a), is a generalized diagonal that starts at vertex A and terminates at vertex D after three collisions; hence, it has length three. A similar shift of P_0 to vertex A for the other three types of period-ten orbits, shown in Figure 7(b)–7(d), leads to different generalized diagonals that start at A with slopes $2/3$, $3/2$ and 4, respectively; these are shown in Figure 8(b)–8(d), respectively. (We chose a somewhat peculiar labelling of the panels in Figure 8 such that the labels of the first two columns match the labels of the panels in Figure 7.) Note that each of these generalized diagonals also has length three. Analogous to Definition 2.2 of equivalent periodic orbits, the four generalized diagonals in Figure 8(a)–8(d) are from different equivalence classes, because they have respectively 0, 1, 2 and 3 of their three collision points on the horizontal sides of the square billiard.

EXAMPLE 3.18. Figures 8(e) and 8(f) show the generalized diagonals obtained when we use the approach from Example 3.17 for the period-six orbits of types $(2, 4)$ and $(4, 2)$ in Figures 2(c) and 2(a), respectively. These generalized diagonals each have length one and have zero or one collision point on the horizontal sides of the square billiard.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

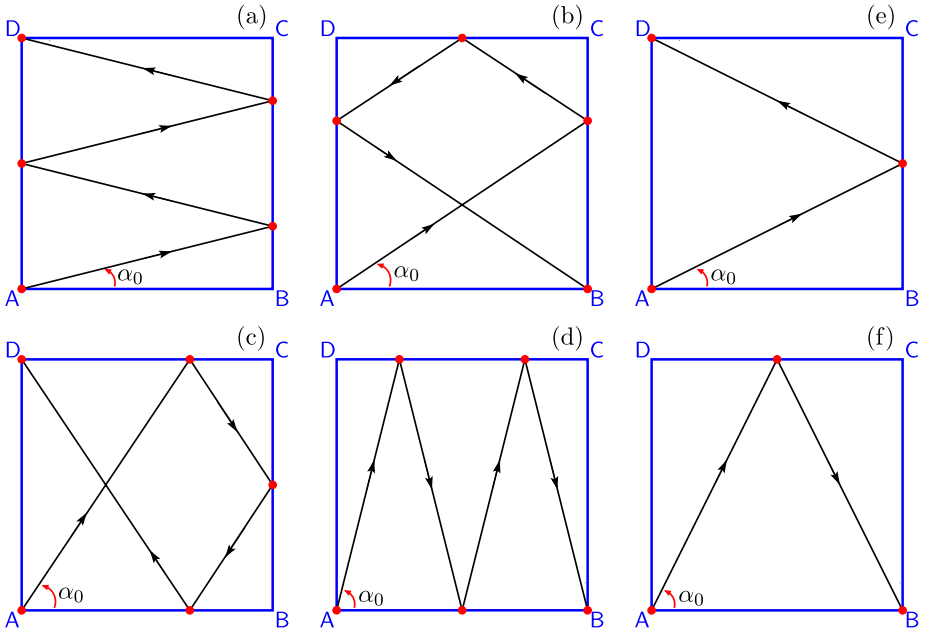

FIGURE 8. Examples of generalized diagonals. The panels in the first two columns, labelled (a)–(d), show the four different types of generalized diagonals of length three, and the right-most two panels labelled (e) and (f) show the two types of length two; compare also with Figure 7.

PROPOSITION 3.19. *Consider a generalized diagonal for the square billiard* ABCD *that starts at vertex* A *with rational slope* m/n*, where* $m, n \in \mathbb{N}$ *such that* $\gcd(m, n) = 1$*. Then, this generalized diagonal has length* $m + n - 2$ *given by* $m - 1$ *collision points on the horizontal sides* AB *or* CD *and* $n - 1$ *collision points on the vertical sides* BC *or* DA.

PROOF. The unfolded trajectory of such a generalized diagonal is the line through the origin with slope m/n that passes through all vertex points $(i, j) \in \mathbb{Z} \times \mathbb{Z}$ that are integer multiples of (n, m). There are no other points $(i, j) \in \mathbb{Z} \times \mathbb{Z}$ on this line, because $\gcd(m, n) = 1$. From $(0, 0)$ to (n, m), we encounter exactly $m - 1$ vertical reflections and exactly $n - 1$ horizontal reflections, which means that there are $m - 1$ collision points on the horizontal sides AB or CD and $n - 1$ collision points on the vertical sides BC or DA. The length of this singular trajectory is, thus, $m - 1 + n - 1 = m + n - 2$, as required. □

Observe that a generalized diagonal can never start and terminate at the same vertex. Indeed, if a generalized diagonal for the square ABCD starts, for example, at vertex A, then it can only terminate at A if the unfolded trajectory is a line that passes through one of the points $(i, j) \in \mathbb{Z} \times \mathbb{Z}$ with both i and j even; we may assume $\gcd(i, j) = 2$, because otherwise, (i, j) is not the first vertex at which the generalized diagonal terminates.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

The unfolded trajectory is then a line with slope $j/i = (j/2)/(i/2)$, which passes through the vertex point $(i/2, j/2)$ before reaching A. Since $\gcd(i,j) = 2$, either $i/2$ or $j/2$ or both will be odd, which means that the generalized diagonal already terminated in the vertex B, D or C, respectively.

REMARK 3.20. Proposition 3.19, combined with Proposition 3.8, enables us to list all generalized diagonals of a given length that start at vertex A. In total, there are twice the number of different generalized diagonals, because we consider a billiard trajectory equal to its reversed-direction copy. More precisely, there are exactly two generalized diagonals in each equivalence class, namely, a generalized diagonal that starts at vertex A and terminates at one of the other vertices, and a reflected copy of this generalized diagonal that starts and terminates at the other (remaining) two vertices.

Figures 2, 7 and 8 suggest a relation between the equivalence class of a periodic orbit in the square billiard of a particular period, and the types and lengths of the generalized diagonals that bound the families in this class. Indeed, we find that all families of periodic orbits of a given type are related to the same generalized diagonal. More precisely, we have the following result.

THEOREM 3.21. *Given $p, q \in \mathbb{N}$ such that $\gcd(p,q) = 2$, consider the billiard trajectory for the square billiard generated by a pair $\langle P_0, \alpha_0 \rangle$ with $\tan(\alpha_0) = p/q$ fixed and $P_0 \in [0,1)$ varying. Then, the billiard trajectory is one of the following:*

$$\begin{cases} \text{a singular orbit}, & \text{if } P_0 = \dfrac{2\ell}{p} \text{ for } \ell = 0, 1, \ldots, \dfrac{p}{2} - 1, \\ \text{a periodic orbit in } C_{p+q}(p), & \text{otherwise}, \end{cases}$$

and the singular orbit lies on a generalized diagonal of length $(p+q)/2 - 2$ with precisely $p/2 - 1$ collision points on the horizontal sides AB *or* CD *and $q/2 - 1$ collision points on the vertical sides* BC *or* DA.

PROOF. Since $\tan(\alpha_0) = p/q$ is rational, this billiard trajectory cannot be a nonperiodic orbit, because nonperiodic orbits unfold to lines with irrational slopes; see Remark 3.3. If it is periodic, then it is of type (p,q) and Definitions 2.2 and 3.6 imply that such a periodic orbit will be a member of the class $C_K(p)$, with $K = p + q$. Furthermore, the billiard trajectory unfolds to a line (or line segment) with fixed slope p/q, so if it is singular, it must lie on a generalized diagonal with this slope. Since $\gcd(p,q) = 2$, Proposition 3.19 implies that the generalized diagonal will have length $p/2 + q/2 - 2$ as required, and it will, indeed, have $p/2 - 1$ collision points on the horizontal sides AB or CD and the remaining $q/2 - 1$ collision points on the vertical sides BC or DA. Hence, the proof is complete as soon as we show that a billiard trajectory generated by the pair $\langle P_0, \alpha_0 \rangle$ is singular if and only if $P_0 = 2\ell/p$ for $\ell = 0, 1, \ldots, p/2 - 1$.

Consider the unfolding of a generalized diagonal that starts at vertex A with slope $m/n = p/q$. This generalized diagonal terminates at the point $(n,m) = (q/2, p/2) \in \mathbb{Z} \times \mathbb{Z}$, because $\gcd(p,q) = 2$, so $m/n = p/q$ and $\gcd(m,n) = 1$. The $p/2 - 1$ collision

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

points on the horizontal sides AB or CD are given by the $m-1$ intersection points with the horizontal lines $\{y=j\}$ for $j=1,\ldots,m-1$. If we translate this line down over j integer units, and left over another, say, i integer units for some $0 \le i < n$, we can move any one of these intersection points to the segment $[0,1]$ on the x-axis, which corresponds to the original side AB of the square billiard. The translated points still lie on a generalized diagonal, because both the start and terminal vertex points at $(0,0)$ and (n,m) map to $(-i,-j)$ and $(n-i,\ m-j)$, respectively, which are also vertex points. The corresponding values for P_0 are then given by the distances to the origin of the translated intersection points, or equivalently, the distances to the vertices (i,j) for the intersection points on the lines $\{y=j\}$, with $j=1,\ldots,m-1$. Note that there can be no other values for $P_0 \in [0,1)$ that lead to singular orbits, because their unfolded trajectories must lie on generalized diagonals with slope $m/n = p/q$ that have at most $m-1$ collision points on the side AB. Hence, if we include $P_0 = 0$, there are only m candidates. The intersection points with integer y-coordinates $y=j$ have x-coordinates $x = jn/m$, so the values for P_0 that lead to singular orbits are

$$P_0 = (jn/m) \pmod 1 \in [0,1) \quad \text{for } j = 0,\ldots,m-1.$$

We claim that this set of points is the same as the set of points $P_0 = 2\ell/p$ for $\ell = 0,1,\ldots,p/2-1 = m-1$. To see this, first note that the points $(j-1)n/m$ and jn/m differ by n/m for all $j=1,\ldots,m-1$; we say that they are uniformly distributed on the interval $[0,n]$. It is important to realize that the first point is located at 0. Imagine wrapping this interval n times around a circle with circumference 1. Then, each point jn/m will map to a point on this circle at an arclength from 0 that cannot exceed 1; each point jn/m maps to a different point on the circle, because $\gcd(m,n)=1$. We can view such points as angles $2\pi\theta_j$, measured in radians, with $\theta_j \in [0,1)$. The difference n/m between two neighbouring points $(j-1)n/m$ and jn/m translates on the circle to the arclength distance $(n/m) \pmod 1$, or a rotation by angle $(2\pi n/m) \pmod{2\pi}$, but the points may not be direct neighbours any longer; in other words, the sequence of values θ_j for $j=0,\ldots,m-1$ is not necessarily in increasing order. Next, observe that $(mjn/m) \in \mathbb{Z}$, so $(mjn/m) \pmod 1 = 0$ for all $j=0,\ldots,m-1$. In the complex plane, this means that each point $(jn/m) \pmod 1$ is a solution to the equation

$$(e^{2\pi i\theta})^m = 1,$$

which are determined by the m roots of unity in the complex plane, and these are uniformly distributed on the unit circle. Since $e^0 = 1$, the roots are given by the angles $2\pi\theta_\ell$, with $\theta_\ell = \ell/m = \ell(1/m) \in [0,1)$ for $\ell = 0,\ldots,m-1$ also uniformly distributed on the unit interval $[0,1)$. Therefore, the (unordered) set $\{\theta_j = (jn/m) \pmod 1 \mid j = 0,\ldots,m-1\}$ is the same as the (ordered) set $\{\theta_\ell = \ell/m \mid \ell = 0,\ldots,m-1\}$, because both sets contain exactly m uniformly distributed points in $[0,1)$, starting from $\theta_0 = 0$.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

For each $j = 1, \dots, m-1$, we obtain $\theta_j = (jn/m) \pmod 1$ from the ordered set $\{\theta_\ell = \ell/m \mid \ell = 0, \dots, m-1\}$ by taking the nth neighbour after the point θ_ℓ that corresponds to θ_{j-1}; here, we treat $\theta_0 = 0$ as a neighbour of $\theta_{m-1} = (m-1)/m$. □

REMARK 3.22. The expert reader will recognize the sequence θ_j as a trajectory of the circle map on the unit interval defined by the rigid rotation $x \mapsto x + \omega$ with winding number $\omega = n/m$.

We illustrate Theorem 3.21 with the following example.

EXAMPLE 3.23. Consider the period-ten orbit from Figure 7(c), which is a member of the class $C_{10}(6)$ and generated by the pair $\langle 0.15, \tan^{-1}(3/2)\rangle$. Its corresponding generalized diagonal of length three is shown in Figure 8(c). Figure 9 shows the unfoldings of these two billiard trajectories as two lines with slope $3/2$: the generalized diagonal is the (red) line starting at vertex A at the origin (red point), and the periodic orbit is the (black) line starting at the point $(0.15, 0)$ (black). Here, we used a (light-blue) shading for the tiles in $\mathbb{R}^2$ with orientation ABCD, instead of labelling each reflected vertex. The generalized diagonal has three collision points, which are consecutively located at the points $(2/3, 1)$ (green), $(1, 3/2)$ (light-blue) and $(4/3, 2)$ (green) in the plane. The first and last (green) points are collisions with the sides AB or CD and they determine the possible initial points P_0 on AB that lead to a period-ten orbit when the initial slope of the billiard trajectory is $3/2$. The initial points that are excluded lie on the singular orbits (red lines) obtained by translation of the generalized diagonal through A such that each of the two collision points $(2/3, 1)$, $(4/3, 2)$ are mapped to the interval $[0, 1)$ on the x-axis. For $(2/3, 1)$, this is achieved by starting the generalized diagonal from the vertex $(0, -1)$, which crosses the x-axis at $(2/3, 0)$, and for $(4/3, 2)$, the generalized diagonal should start from vertex point $(-1, -2)$ so that it crosses the x-axis at the point $(1/3, 0)$. Hence, all period-ten orbits in the class $C_{10}(6)$ are generated by pairs $\langle P_0, \tan^{-1}(3/2)\rangle$ with $P_0 \in (0, 1/3)$, $P_0 \in (1/3, 2/3)$ or $P_0 \in (2/3, 1)$, while starting points $P_0 = 1/3$ and $P_0 = 2/3$ generate singular orbits with that angle. There are no other period-ten orbits that collide six times with the horizontal and four times with the vertical sides of the square ABCD.

REMARK 3.24. Note that each period-ten orbit in the class $C_{10}(6)$ has three collision points on the side AB. Each such period-ten orbit is uniquely identified by the generator pair $\langle P_0, \tan^{-1}(3/2)\rangle$ with $P_0 \in (0, 1/3)$. Indeed, any period-ten orbit generated by the pair $\langle P, \tan^{-1}(3/2)\rangle$ with $P \in (2/3, 1)$ is exactly the same period-ten orbit as the one starting from $P_0 = P - 2/3 \in (0, 1/3)$, which encounters $P \in (2/3, 1)$ as its sixth collision point; see Figure 9. Similarly, any period-ten orbit that starts from $P \in (1/3, 2/3)$ is the same as the one starting from $P_0 = 2/3 - P$, which encounters the point $P \in (1/3, 2/3)$ on AB after three collisions; this is shown in Figure 9 on an (unshaded) tile with the vertically reflected orientation BCDA. Hence, the family of period-ten orbits in the class $C_{10}(6)$ is completely represented by an initial point P_0 up to distance $1/3$ from A along the side AB and the outgoing line with slope $3/2$.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

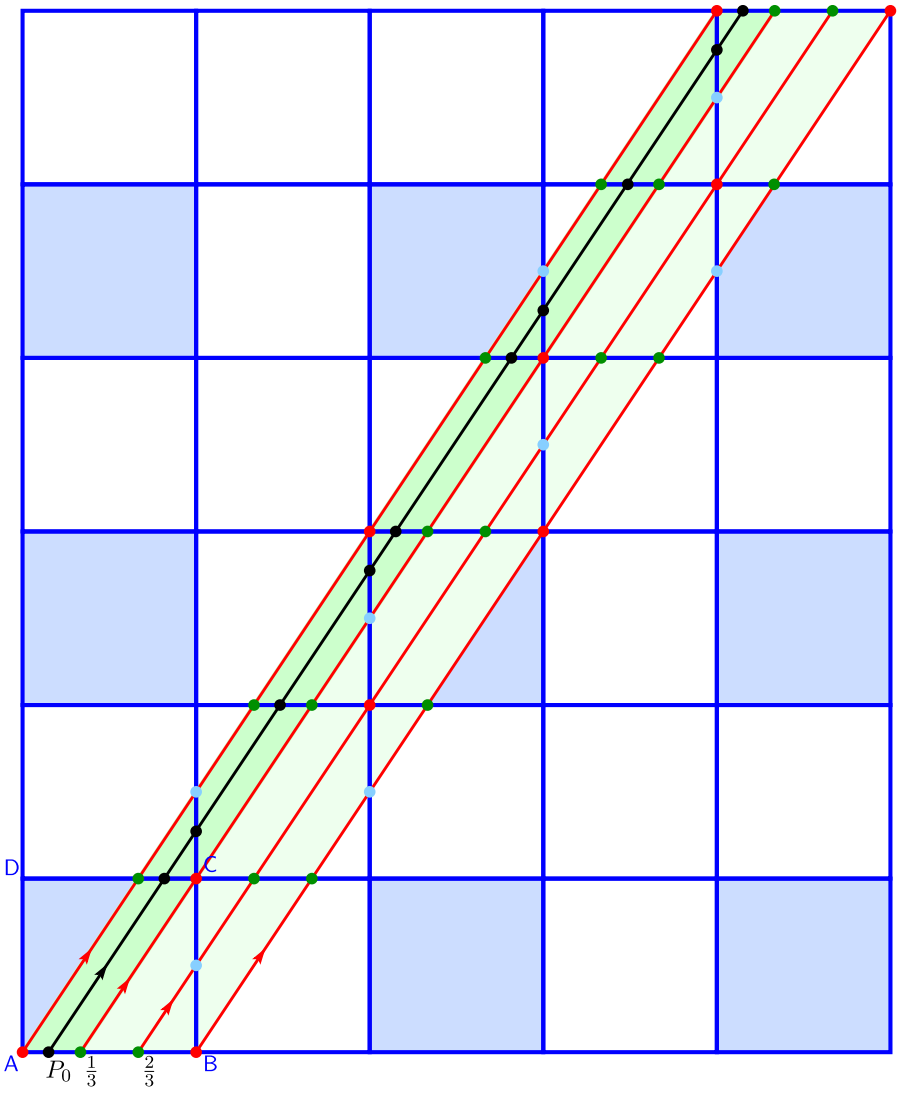

FIGURE 9. Unfolding of all possible period-ten orbits of type $(6, 4)$, and associated bounding singular orbits that start on the side AB of the square ABCD. Shown are the period-ten orbit (black) from Figure 7(c) that starts at $P_0 = 0.15$ and the generalized diagonal (red) from Figure 8(c) starting at vertex A, together with three translated versions that give the two singular orbits starting at $1/3$, $2/3$ and the generalized diagonal starting at vertex B. The dark (green) shaded strip represents the entire family in $C_{10}(6)$, while the light (green) shaded strip indicates the regime of existence starting from any point on the side AB; the (light-blue) shaded tiles correspond to tables with the orientation ABCD.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

4. Rectangular billiards

So far, we restricted attention to the square billiard, that is, a rectangular table with aspect ratio 1 : 1. In this section, we extend our results to rectangular tables with arbitrary aspect ratios. More precisely, we consider a rectangular table represented by the scaled table ABCD with sides AB and CD of length 1 and sides BC and DA of length ϱ, where $0 < \varrho < \infty$. Recall that, just as for the square billiard, a constant scaling of all sides of the table does not change the number or types of periodic orbits. The question addressed in this section is whether a scaling of just the sides BC and DA changes the number or types of periodic orbits.

REMARK 4.1. For the square table, it is straightforward to rotate the table and place any of the four vertices at the origin. For the rectangular table with aspect ratio $1 : \varrho$, this holds for vertices A and C only; the quarter rotations that place either vertex B or vertex D at the origin result in a scaled table with aspect ratio $1 : (1/\varrho)$. This is the reason why we made a distinction between periodic orbits of type (p, q) and those of type (q, p), or equivalently, between the equivalence classes $C_K(p)$ and $C_K(q) = C_K(K - p)$.

The rectangular table is a natural and perhaps most straightforward extension from the square table, because it is the generic linear transformation of the square table that preserves the angles between all four sides. Consequently, the technique of unfolding a billiard trajectory to a straight line in the plane $\mathbb{R}^2$ can be applied in the same way as for the square billiard; essentially, the unfolding takes place relative to the vertices that now lie on the transformed mesh $\mathbb{Z} \times \varrho\,\mathbb{Z}$, rather than the original square mesh $\mathbb{Z} \times \mathbb{Z}$. This also means that there exists a one-to-one correspondence between billiard trajectories on the square billiard and those on the rectangular billiard. More precisely, any billiard trajectory on the square, be it singular, periodic or nonperiodic, can be unfolded and then mapped via the appropriate linear transformation to a straight line on the transformed mesh $\mathbb{Z} \times \varrho\,\mathbb{Z}$ that again corresponds to a singular, periodic or nonperiodic orbit on the rectangular billiard, respectively. Here, the slope of the unfolded trajectory will change (by a factor ϱ), but the line is positioned the same way relative to the vertices.

EXAMPLE 4.2. Compare the period-six orbit from Figures 2(c) and 3 with the period-six orbit shown in Figure 10 that lies in the rectangle ABCD with aspect ratio $1 : \varrho$, where $\varrho = 3/2$. As in Figure 3, the trajectory has been unfolded on the plane as well. Observe that the sides BC and DA of the rectangular billiard are 1.5 times longer than its sides AB and CD. Hence, the line that connects the point $P_0 = (0.7, 0)$ with its translated copy after a total of two vertical and four horizontal reflections has slope $3/4 = 2\varrho/4 = \varrho/2$; hence, the slope of the unfolded trajectory on the rectangular table changes by a factor ϱ compared with the slope $1/2$ for the line in Figure 3 associated with the square table.

Other than the adjustment by this factor ϱ, the results for the square billiard presented in the previous section naturally extend to the rectangular billiard with

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

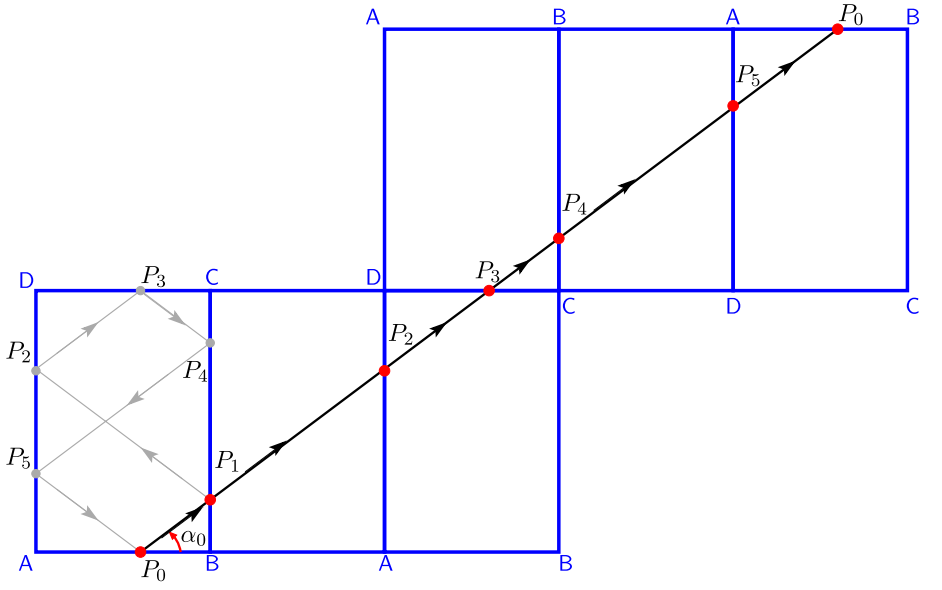

FIGURE 10. Unfolding of a period-six orbit of type $(2, 4)$ in the rectangular billiard ABCD with aspect ratio $1 : (3/2)$; compare with Figures 2(c) and 3.

aspect ratio $1 : \varrho$. We briefly summarize these extended results here; note that the characterization includes the square billiard as the case $\varrho = 1$.

4.1. Classification of periodic orbits for the rectangular billiard The rectangular billiard has the same equivalence classes of periodic orbits, which can be constructed in the same way after adjusting the slope as required for the given aspect ratio. In particular, all periodic orbits for the rectangular billiard have even period, and the product of the slope of the trajectory and aspect ratio of the rectangle is rational; see also Theorem 3.5. More precisely, we have the following extended version of Theorem 3.21.

PROPOSITION 4.3. *For the rectangular billiard* ABCD *with aspect ratio* $1 : \varrho$*, there exists a periodic orbit that has* p *distinct collision points on the sides* AB *or* CD *and* q *distinct collision points on the sides* BC *or* DA *for any* $p, q \in \mathbb{N}$ *with* $\gcd(p, q) = 2$*. This periodic orbit is from the equivalence class* $C_K(p)$ *with* $K = p + q$*, and it is generated by a pair* $\langle P_0, \alpha_0 \rangle$*, with* $\alpha_0 \in (0, \pi/2]$ *such that* $\tan(\alpha_0) = \varrho p/q$*. The point* P_0 *can take almost any value in the interval* $[0, 1)$*; the exceptions given by*

$$P_0 = \frac{2\ell}{p} \quad \text{for } \ell = 0, 1, \ldots, \frac{p}{2} - 1,$$

lead to a singular orbit, which lies on a generalized diagonal of length $(p + q)/2 - 2$ *with precisely* $p/2 - 1$ *collision points on the horizontal sides* AB *or* CD *and* $q/2 - 1$ *collision points on the vertical sides* BC *or* DA.

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

Given $K = 2N$ for some $N \in \mathbb{N}$, we can construct period-K orbits of type (p, q) for any $p, q \in \mathbb{N}$ such that $K = p + q$ and $\gcd(p, q) = 2$. Hence, just as for the square billiard, we can apply Proposition 3.8 which means that Corollary 3.9 also holds for the rectangular billiard: there are $\varphi(N)$ unique combinations for p and q, leading to $\varphi(N)$ different types of period-K orbits, generated by $\varphi(N)$ different angles in the interval $(0, \pi/2]$, which represent a total of $\varphi(N)$ different equivalence classes.

REMARK 4.4. As for the square billiard, any rectangle will have periodic orbits of all (even) periods, including the periodic orbits with period $K = 2$. Note that the period-two orbits are not covered in Proposition 4.3, because either $p = 0$ or $q = 0$ when $K = 2$. As can be inferred from Definition 2.2, there are again two different equivalence classes, namely, $C_2(2)$ and $C_2(0)$.

The initial angle α_0 is only unique for period-two orbits; its value will vary for other equivalence classes, because of the dependence on the aspect ratio of the billiard table. The converse is also true.

EXAMPLE 4.5. The angle α_0 with $\tan(\alpha_0) = 1/2$ generates a period-six orbit for the square billiard table, but this will be a period-ten orbit on the rectangle with aspect ratio $1 : 2$, while it is a period-four orbit on the rectangle with aspect ratio $1 : (1/2)$, or equivalently, by starting a billiard trajectory with this slope on a vertical side of length 2.

EXAMPLE 4.6. Similarly, the angle α_0 with $\tan(\alpha_0) = 1/2$ generates a period-six orbit for the square billiard table, but this same angle produces periodic orbits of periods ten or 22 on a rectangle with aspect ratio $1 : (3/4)$ when starting from a horizontal or a vertical side, respectively; the period-ten orbit is of type $(4, 6)$, in contrast to the period-ten orbit produced for this angle on a table with aspect ratio $1 : 2$, which will be of type $(2, 8)$.

Proposition 3.14 and Corollary 3.15 hold for rectangular billiards as well. More precisely, we have the following properties.

PROPOSITION 4.7. *For the rectangular billiard* ABCD *with aspect ratio* $1 : \varrho$, *consider a billiard trajectory that starts at vertex* A, B, C *or* D. *Define* $\sigma \in \mathbb{R}$ *as the slope of the billiard trajectory measured relative to the side* AB.

- *If* σ/ϱ *is irrational, then this billiard trajectory never collides with another vertex.*
- *If* σ/ϱ *is rational, then this billiard trajectory will always end up in a vertex, that is, it is a singular orbit. Furthermore, the trajectory generated in the opposite direction, using the reflected (negative) slope, will also be singular.*

Note that it is necessary to specify the side with respect to which the slope is measured; if the slope is σ when measured with respect to the side AB, then it is also σ when measured with respect to the side CD, but it will be $1/\sigma$ when measured with

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

respect to the sides BC or DA. For the square table, this difference does not matter, because $1/\sigma$ is (ir)rational if and only if σ is (ir)rational. However, for the rectangular billiard, the adjustment by the factor ϱ does not necessarily preserve this equivalence if ϱ is itself irrational. More precisely, if the slope is σ when measured with respect to, say, the side BC, then we effectively consider the rotated table with aspect ratio $1 : (1/\varrho)$, and the behaviour of the billiard trajectory starting at vertex B will be determined by whether $(1/\sigma)/\varrho = 1/\varrho\sigma$ is irrational or not; this is not equivalent to asking whether σ/ϱ is irrational or not.

5. Discussion

We completely classified the existence and nature of all periodic orbits for a rectangular billiard table with aspect ratio $1 : \varrho$, where $0 < \varrho < \infty$; we discussed the square billiard table in full detail, which is included in this setting as the case with $\varrho = 1$. The class of rectangular billiards is special, because it is the only class of polygonal billiards for which the classification of all periodic orbits is preserved under a nonuniform scaling [13, 21]. The special property that the sides of a rectangular table are perpendicular to each other enables us to characterize the periodic orbits for this class of billiard tables in unprecedented detail. In particular, we show that rectangular billiards admit period-K orbits for any even period, but not for any odd period $K \in \mathbb{N}$. As soon as $K \geq 6$, there are different types of period-K orbits, determined by the difference between the numbers of collisions with the two pairs of parallel sides. Each type is completely determined by the slope that the periodic orbit should have with respect to one of the sides of the rectangle. Furthermore, the periodic orbit can be realized with any initial point on this side, except for a finite set of exactly $K/2 - 1$ points, other than the table corners, that give rise to singular orbits; the unfolding of such singular orbits leads to generalized diagonals in the plane. We defined equivalence classes for each type and proved that the total number of equivalence classes for period-K orbits, with $K \in 2\,\mathbb{N}$, is given by Euler's totient function $\varphi(N)$ evaluated at $N = K/2$; the totients $n \in \{1, \ldots, N\}$ such that $\gcd(n, N) = 1$ define pairs (m, n) with $m = N - n$ that uniquely define the slopes needed to generate a periodic orbit of a specific type.

Interestingly, Euler's totient function also appears in the theory of so-called Sturmian sequences [2]. A Sturmian sequence is formed by an infinitely long series of 0s and 1s that never repeats. The definition involves particular restrictions that we will not explain here, but it suffices to say that any Sturmian sequence can be viewed as a nonperiodic orbit in the square billiard by recording its collisions with the vertical and horizontal sides of the billiard, labelling them 0 and 1, respectively, or vice versa. Mignosi [40] found an explicit formula for the number of *factors* (different subsequences) of a particular length m in *any* Sturmian sequence; this formula includes a (weighted) summation of Euler's totient function evaluated over all integers from 1 to m. Our result suggests a relation between the different types of periodic orbits with period $K = 2\,m$ and the *specific* Sturmian sequence or sequences that correspond to

https://doi.org/10.1017/S1446181125100163 Published online by Cambridge University Press

nonperiodic orbits lying, in some sense, near a period-K orbit. Indeed, Mignosi's result was proven again in [7] using geometric arguments reminiscent of our Figure 9. The investigation of this relation is left for further work.

As mentioned, the difference between the square and rectangular billiard is a stretching of one pair of parallel sides, which leads to an adjustment by the aspect-ratio parameter ϱ of the slopes required to generate particular periodic orbits. It is tempting to apply a more general transformation, for example, one that scales and shears the sides, such that the billiard table becomes a parallelogram. Unfortunately, the unfolding technique generally fails to tile the plane for parallelogram billiards [23, 52, 56]. We find that there is very little known regarding parallelogram billiards. Do parallelograms also only admit periodic orbits with even periods? What are necessary and sufficient conditions for the existence of a period-K orbit and how do we construct such a periodic orbit for the parallelogram?

The nonexistence of periodic orbits with odd periods has been proven for parallelograms consisting of two equilateral triangles glued together, that is, one angle is $\pi/3$ radians [3]. Furthermore, it is known that periodic orbits for the parallelogram with angles $\pi/4$ and $3\pi/4$ do not persist under perturbations, that is, so-called stable periodic orbits do not exist for this parallelogram table [45]. More generally, it is known that polygonal billiard tables with angles equal to rational multiples of π have many periodic trajectories [10, 18, 39, 52]. However, almost nothing is known for billiard tables with irrational angles. For example, as we already alluded to in the introduction, it is an open question whether periodic orbits exist for triangular billiard tables with arbitrary (irrational) angles. The (rational) angles for acute triangles that lead to a *lattice polygon*, that is, triangles that unfold to a planar tiling, have all been listed [36, 44]. Hooper [29] has shown that periodic orbits for right-triangle billiards are never stable. In contrast, if one perturbs the right triangle such that it has an obtuse angle, then it will always have a stable periodic orbit [47, 48]; the proof is restricted to obtuse angles that do not exceed $5\pi/8$, which is 112.5 degrees, but their computer-assisted approach stops at an angle of 100 degrees. The current world record is a different computer-assisted proof that every obtuse triangle with obtuse angle at most 112.3 degrees has a periodic orbit [53]. For larger angles, we are only aware of results of periodic orbits for isosceles triangles [30]. We believe that progress can be made from the study of periodic orbits in a parallelogram table, which we view as continuous deformations from corresponding periodic orbits in the rectangular table. Results in this direction are left for future work.

Acknowledgments

Hongjia Chen is grateful to Sean Gasiorek for many fruitful discussions regarding billiards. Hinke Osinga thanks Moon Duchin for her directions to the more recent literature on the topic and Sergei Tabachnikov for his thorough reading of, and thoughtful comments on an earlier draft of this paper.

References

[1] V. I. Arnold, *Mathematical methods of classical mechanics*, Volume 60 of *Grad. Texts in Math. (GTM)* (Springer, Berlin, 1989); doi:10.1007/978-1-4757-2063-1.

[2] P. Arnoux, "Sturmian sequences", in: *Substitutions in dynamics, arithmetics and combinatorics*, Volume 1794 of *Lecture Notes in Math. (LNM)* (eds. N. P. Fogg, V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel) (Springer, Berlin-Heidelberg, 2002) 143–198; doi:10.1007/3-540-45714-3_6.

[3] B. R. Baer, F. Gilani, Z. Han and R. Umble, "Periodic orbits on obtuse edge tessellating polygons", *Pi Mu Epsilon J.* **15**(6) (2022) 321–331; available at https://arxiv.org/abs/1911.01397.

[4] M. Baker, "Alhazen's problem", *Amer. J. Math.* **4** (1881) 327–331; doi:10.2307/2369168.

[5] A. M. Baxter and R. Umble, "Periodic orbits for billiards on an equilateral triangle", *Amer. Math. Monthly* **115**(6) (2008) 479–491; doi:10.1080/00029890.2008.11920555.

[6] M. V. Berry, "Quantum scars of classical closed orbits in phase space", *Proc. Roy. Soc. Edinburgh Sect. A* **423**(1864) (1989) 219–231; doi:10.1098/rspa.1989.0052.

[7] J. Berstel and M. Pocciola, "A geometric proof of the enumeration formula for Sturmian words", *Internat. J. Algebra Comput.* **3** (1993) 349–355; doi:10.1142/S0218196793000238.

[8] M. Bialy, C. Fierobe, A. Glutsyuk, M. Levi, A. Plakhov and S. Tabachnikov, "Open problems on billiards and geometric optics", *Arnold Math. J.* **8**(3–4) (2022) 411–422; doi:10.1007/s40598-022-00198-y.

[9] C. Boldrighini, M. Keane and F. Marchetti, "Billiards in polygons", *Ann. Probab.* **6** (1978) 532–540; doi:10.1214/aop/1176995475.

[10] M. Boshernitzan, G. Galperin, T. Krüger and S. Troubetzkoy, "Periodic billiard orbits are dense in rational polygons", *Trans. Amer. Math. Soc.* **350** (1998) 3523–3535; doi:10.1090/S0002-9947-98-02089-3.

[11] L. A. Bunimovich and C. P. Dettmann, "Open circular billiards and the Riemann hypothesis", *Phys. Rev. Lett.* **94** (2005) article ID: 100201; doi:10.1103/PhysRevLett.94.100201.

[12] L. A. Bunimovich, et al., *Dynamical systems, ergodic theory and applications*, Volume 100 of *Encyclopedia Math. Sci. Ser.* (ed. Y. G. Sinai) (Springer, Berlin–Heidelberg, 2000).

[13] A. Calderon, S. Coles, D. Davis, J. Lanier and A. Oliveira, "How to hear the shape of a billiard table", Preprint, 2018, arXiv:1806.09644.

[14] Y. F. Chen K. F. Huang and Y. P. Lan, "Localization of wave patterns on classical periodic orbits in a square billiard", *Phys. Rev. E* **66**(3) (2002) article ID: 046215; doi:10.1103/PhysRevE.66.046215.

[15] N. Chernov and R. Markarian, *Chaotic billiards*, Volume 127 of *Math. Surveys Monogr. (MSM)* (American Mathematical Society, Providence, RI, 2006).

[16] H. S. M. Coxeter and S. L. Greitzer, *Geometry revisited* (Mathematical Association of America, Washington, DC, 1967).

[17] D. Davis, "Lines in positive genus: an introduction to flat surfaces", in: *Dynamics done with your bare hands*, Volume 26 of *EMS Ser. Lect. Math.* (eds. F. Dal'Bo, F. Ledrappier and A. Wilkinson) (European Mathematical Society, Zürich, 2016) 1–55; doi:10.4171/168-1/1.

[18] L. DeMarco, "The conformal geometry of billiards", *Bull. Amer. Math. Soc. (N.S.)* **48** (2011) 33–52; doi:10.1090/S0273-0979-2010-01322-7.

[19] C. P. Dettmann, "Recent advances in open billiards with some open problems", in: *Frontiers in the study of chaotic dynamical systems with open problems*, Volume 16 of *World Sci. Ser. Nonlinear Sci. Ser. B* (eds. E. Zeraoulia and J. C. Sprott) (World Scientific, Singapore, 2011) 195–218; doi:10.1142/9789814340700_0011.

[20] C. P. Dettmann, "Diffusion in the Lorentz gas", *Comm. Theoret. Phys.* **62** (2014) 521–540; doi:10.1088/0253-6102/62/4/10.

[21] M. Duchin, V. Erlandsson, C. J. Leininger and C. Sadanand, "You can hear the shape of a billiard table: symbolic dynamics and rigidity for flat surfaces", *Comment. Math. Helv.* **96** (2019) 421–463; doi:10.4171/cmh/516.

[22] P. Dulio and A. Frosini, "Integer orbits in rectangular lattice billiards", *Discrete Appl. Math.* **358** (2024) 302–319; doi:10.1016/j.dam.2024.07.019.

[23] R. H. Fox and R. B. Kershner, "Concerning the transitive properties of geodesics on a rational polyhedron", *Duke Math. J.* **2** (1936) 147–150; doi:10.1215/S0012-7094-36-00213-2.
[24] E. Fredkin and T. Toffoli, "Conservative logic", *Internat. J. Theoret. Phys.* **21** (1982) 219–253; doi:10.1007/BF01857727.
[25] G. Galperin, "Playing pool with π (the number π from a billiard point of view)", *Regul. Chaotic Dyn.* **8** (2003) 375–394; doi:10.1070/RD2003v008n04ABEH000252.
[26] Z. Ge, et al., "Direct visualization of relativistic quantum scars in graphene quantum dots", *Nature* **635** (2024) 841–846; doi:10.1038/s41586-024-08190-6.
[27] E. Gutkin, "Billiard dynamics: an updated survey with the emphasis on open problems", *Chaos* **22** (2012) article ID: 026116; doi:10.1063/1.4729307.
[28] G. T. Herman and A. Kuba, *Advances in discrete tomography and its applications* (Birkhäuser, Boston, MA, 2007); doi:10.1007/978-0-8176-4543-4.
[29] W. P. Hooper, "Periodic billiard paths in right triangles are unstable", *Geom. Dedicata* **125** (2007) 39–46; doi:10.1007/s10711-007-9129-9.
[30] W. P. Hooper and R. E. Schwartz, "Billiards in nearly isosceles triangles", *J. Mod. Dyn.* **3** (2009) 159–231; doi:10.3934/jmd.2009.3.159.
[31] S. R. Jain and R. Samajdar, "Nodal portraits of quantum billiards: domains, lines, and statistics", *Rev. Modern Phys.* **89** (2017) article ID: 045005; doi:10.1103/RevModPhys.89.045005.
[32] V. Kaloshin and A. Sorrentino, "Inverse problems and rigidity questions in billiard dynamics", *Ergodic Theory Dynam. Systems* **42** (Anatole Katok Memorial Issue Part 2) (2022) 1023–1056; doi:10.1017/etds.2021.37.
[33] A. Katok, "The growth rate for the number of singular and periodic orbits for a polygonal billiard", *Comm. Math. Phys.* **111** (1987) 151–160; doi:10.1007/BF01239021.
[34] A. Katok, *Five most resistant problems in dynamics*, Berkeley: MSRI-Evans Lecture (27 September 2004); available at: https://www.impan.pl/~biuletyn/arch/5problems.pdf (see also https://www.msri.org/workshops/267/schedules/1789; both last accessed 8 September 2025).
[35] A. B. Katok, "Billiard table as a playground for a mathematician", in: *Surveys in modern mathematics*, Volume 321 of *London Math. Soc. Lecture Note Ser.* (eds. V. Prasolov and Yu. Ilyashenko) (Cambridge University Press, Cambridge, 2005) 216–242; doi:10.1017/CBO9780511614156.010.
[36] R. Kenyon and J. Smillie, "Billiards on rational-angled triangles", *Comment. Math. Helv.* **75** (2000) 65–108; doi:10.1007/s000140050113.
[37] N. J. Lennes, "On the motion of a ball on a billiard table", *Amer. Math. Monthly* **12**(8–9) (1905) 152–155; doi:10.1080/00029890.1905.11997250.
[38] P. Loomis, M. Plytage and J. Polhill, "Summing up the Euler φ function", *College Math. J.* **39** (2008) 34–42; doi:10.1080/07468342.2008.11922272.
[39] H. Masur and S. Tabachnikov, "Rational billiards and flat structures", in: *Handbook of dynamical systems*, Volume 1A (eds. B. Hasselblatt and A. Katok) (Elsevier, North Holland, 2002) 1015–1089; doi:10.1016/S1874-575X(02)80015-7.
[40] F. Mignosi, "On the number of factors of Sturmian words", *Theoret. Comput. Sci.* **82** (1991) 71–84; doi:10.1016/0304-3975(91)90172-X.
[41] J. Orchard, F. Frascoli, L. Rondoni and C. Mejía-Monasterio, "Particle transport in open polygonal billiards: a scattering map", *Chaos* **34**(12) (2024) article ID: 123158; doi:10.1063/5.0219730.
[42] E. K. Petersen, *Ergodic theory* (Cambridge University Press, Cambridge, 1983); doi:10.1017/CBO9780511608728.
[43] B. Polster and M. Ross, *Math goes to the movies* (Johns Hopkins University Press, Baltimore, MD, 2012); doi:10.1353/book.16372.
[44] J.-C. Puchta, "On triangular billiards", *Comment. Math. Helv.* **76** (2001) 501–505; doi:10.1007/PL00013215.
[45] U. Rozikov, *An introduction to mathematical billiards* (World Scientific, Singapore, 2018); doi:10.1142/11162.
[46] W. Schramm, "The Fourier transform of functions of the greatest common divisor", *Integers* **8** (2008) article ID: A50; available from https://www.emis.de/journals/INTEGERS/papers/i50/i50.pdf.

[47] R. E. Schwartz, "Obtuse triangular billiards. I. Near the (2, 3, 6) triangle", *Exp. Math.* **15** (2006) 161–182; doi:10.1080/10586458.2006.10128961.
[48] R. E. Schwartz, "Obtuse triangular billiards II: one hundred degrees worth of periodic trajectories", *Exp. Math.* **18** (2009) 137–171; doi:10.1080/10586458.2009.10128891.
[49] Y. G. Sinai, "Hyperbolic billiards", in: *Proc. Int. Cong. Math.*, Kyoto 1990, Volume 1 of *Math. Soc. Japan* (ed. I. Satake) (Springer, 1991) 249–260; available at https://www.mathunion.org/fileadmin/ICM/Proceedings/ICM1990.1/ICM1990.1.ocr.pdf.
[50] Y. G. Sinai, "What is. . .a billiard?", *Notices Amer. Math. Soc.* **51**(4) (2004) 412–413; available at http://www.ams.org/notices/200404/what-is.pdf.
[51] S. Tabachnikov, "Fagnano orbits of polygonal dual billiards", *Geom. Dedicata* **77** (1999) 279–286; doi:10.1023/A:1005199315944.
[52] S. Tabachnikov, *Geometry and billiards*, Volume 30 of *Stud. Math. Libr. (STML)* (American Mathematical Society, Providence, RI, 2005); doi:10.1090/stml/030.
[53] G. Tokarsky, J. Garber, B. Marinov and K. Moore, "One hundred and twelve point three degree theorem", Preprint, 2018, arXiv:1808.06667.
[54] S. Troubetzkoy, "Dual billiards, Fagnano orbits, and regular polygons", *Amer. Math. Monthly* **116** (2009) 251–260; doi:10.1080/00029890.2009.11920934.
[55] W. A. Veech, "Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards", *Invent. Math.* **97** (1989) 553–583; doi:10.1007/BF01388890.
[56] A. N. Zemlyakov and A. B. Katok, 'Topological transitivity of billiards in polygons", *Math. Notes Acad. Sci. USSR* **18** (1975) 760–764; doi:10.1007/BF01818045.