Polygons in billiard orbits

Henk Don

TU Delft, EWI (DIAM), Section Probability Theory, Mekelweg 4, 2628 CD Delft, the Netherlands. Phone: +31 152782546

Abstract

We study the geometry of billiard orbits on rectangular billiards. A truncated billiard orbit induces a partition of the rectangle into polygons. We prove that thirteen is a sharp upper bound for the number of different areas of these polygons.

Keywords: billiard orbit, geometry of partitions
2010 MSC: 11B75

1. Introduction

Let a billiard ball be shot from a corner of a rectangular billiard. Consider the ball as a point, and truncate the orbit somewhere at the boundary. The truncated orbit of the ball generates a partition of the rectangular billiard into polygons, similar to Figure 1. Many of these triangles and quadrangles seem to have the same shape and size. In this paper we will show that (for a fixed shooting angle and stopping point) the number of different areas is at most thirteen. This universal upper bound is the sharpest possible. We also consider rational shooting angles and irrational shooting angles for which the thirteen is never reached.

2. Rotations

The results in this paper are closely related to the Three Gap Theorem (see e.g. [4], [3]) and the Four Gap Theorem (see [1]). The statements of these two theorems are best illustrated by a picture; see Figure 2.

The Three Gap Theorem is naturally associated to the concept of *rotations*. First we recall the theorem and then we discuss rotations on intervals. For $x \in \mathbb{R}$, let $\{x\} = x - \lfloor x \rfloor$ denote its fractional part.

Theorem 1. *(The Three Gap Theorem) Let $n \in \mathbb{N}$ and $\alpha \in (0, 1)$. The numbers*

$$0, \{\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots, \{n\alpha\} \tag{1}$$

Email address: henkdon@gmail.com (Henk Don)

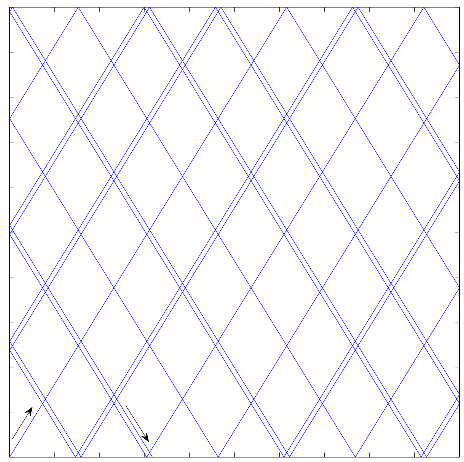

Figure 1: Truncated orbit of a billiard ball. The arrows indicate start and end of the orbit.

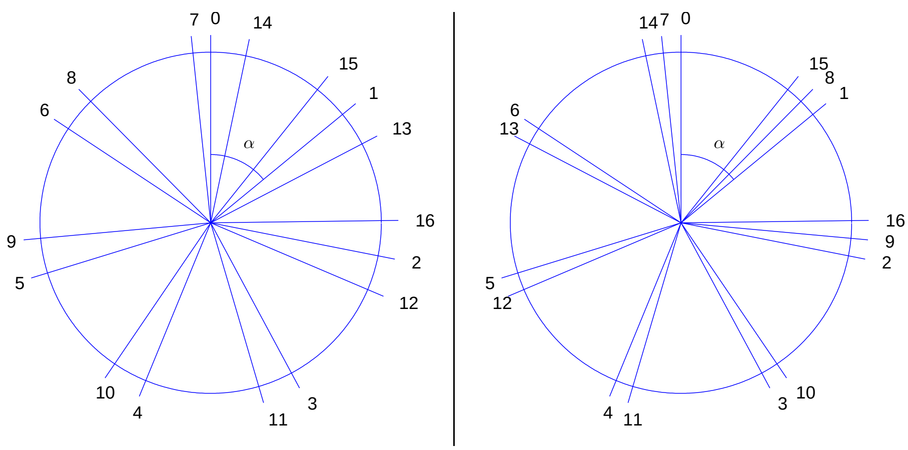

Figure 2: Left figure, the Three Gap Theorem: Cutting a pie n times where each next cut is obtained by shifting the previous one over a fixed angle α gives at most three different sizes of pieces of the pie. Right figure, the Four Gap Theorem: Now the first cut (at 0) works as a 'reflecting boundary'. As soon as it is reached, we continue in the opposite direction. In this case we have after n cuts at most four different sizes.
For this picture we used $\alpha = 0.1405 * 2\pi$ and $n = 17$.

induce a partition of the interval $[0, 1]$ *in subintervals which can have at most three different lenghts. If there are three lengths, then the largest is the sum of the other two.*

Letting $T_\alpha(x) = \{x + \alpha\}$ for $x \in [0, 1]$, the numbers (1) transform into

$$0, T_\alpha(0), T_\alpha^2(0), \ldots, T_\alpha^n(0). \tag{2}$$

If we consider x as a point on the circle of unit circumference, then $T_\alpha(x)$ is obtained by rotating x over a distance α. This gives a more dynamical view of the partition of $[0, 1]$: the partition is induced by a truncated orbit of the rotation map T_α. These observations lead to the following generalization of the Three Gap Theorem:

Property 1. *Let $n_1, n_2 \in \mathbb{N}$, $\alpha \in (0, 1)$ and $a, b \in \mathbb{R}$. The $n_1 + n_2 + 1$ numbers*

$$aT_\alpha^{-n_1}(0) + b, \ldots, aT_\alpha^{-1}(0) + b,\ b\,, aT_\alpha(0) + b, \ldots, aT_\alpha^{n_2}(0) + b \qquad (3)$$

induce a partition of $[b, b + a]$ in subintervals having at most three different lenghts.

This can easily be obtained by taking $n = n_1 + n_2$ in (2), rotating over an appropriate angle and applying the linear map $a \cdot +b$ to the orbit. Actually, a special case of this property already appeared as a theorem in [1]. However, there a complicated proof was given to obtain this result. Vilmos Komornik came up with the idea to place the numbers on the circle, thus obtaining a much simplified and more natural argument [2]. In the sequel we will refer to (3) as a *rotation orbit on* $[b, b + a]$.
There is a slightly stronger property we will need in Remark 1 (see e.g. [1] and [3]):

Property 2. *Take a truncated orbit of a rotation on an interval. Suppose the orbit consists of n numbers. Create another orbit from this by removing the last number. The two partitions induced by these orbits give two sets of lengths. The union of these two sets contains at most three different lengths.*

3. Billiards and the Four Gap Theorem

The billiard in Figure 1 can be seen as a generalization to two dimensions of the pie-cutting process of the Four Gap Theorem, as illustrated in Figure 2. This statement deserves some explanation. Figure 3, a picture in some sense equivalent to the right panel of Figure 2, gives a description of the Four Gap Theorem in terms of a ball bouncing on the unit interval.

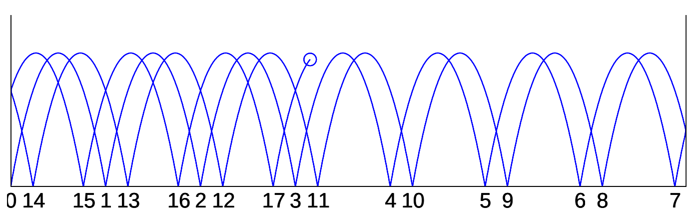

Figure 3: A ball bouncing on an interval between two walls. The Four Gap Theorem makes a statement about the subset of the interval consisting of the landing points of the ball. We used 0.1405 times the length of the interval as bouncing distance.

This figure shows the movement of a ball bouncing between two walls, where we assume that the ball is a point and that there is no loss of energy. The landing

points of the ball build a sequence in the interval. The first n numbers in this sequence (0 included) define a splitting of the interval in n subintervals. The main statement of the Four Gap Theorem is that these subintervals can have at most four different lengths. In Figure 1 we now have a subset of a square, consisting of those points where the billiard ball appears. This observation gives already some reason to consider the billiard as a 2-dimensional generalization of the pie of the Four Gap Theorem. However, we can also argue this point of view in a more mathematical way.
Let $||x||$ denote the distance from x to the nearest integer. For $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, let

$$S_\alpha := (||k\alpha||)_{k=0}^{\infty}.$$

Obviously this is a sequence in $[0, \frac{1}{2}]$. Moreover, it is exactly the sequence of landing points of a ball bouncing between 0 and $\frac{1}{2}$ with horizontal bouncing distance α. The sequence S_α is obtained by 'folding' the sequence of integer multiples of α into the interval $[0, \frac{1}{2}]$. What we mean by this folding is illustrated in Figure 4, where we plot the function $f_1 : [0, \infty) \to [0, \frac{1}{2}]$

$$f_1(x) := ||x||,$$

and illustrate how $[0, \infty)$ is mapped to $[0, \frac{1}{2}]$ by f_1.

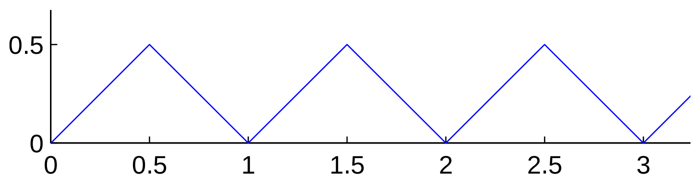

Figure 4: Plot of the folding map $f_1(x) = ||x||$.

Now we concentrate on the billiard: the orbit of the billiard ball is obtained by 'folding' a halfline into a rectangle. Since the shooting angle is arbitrary between 0 and $\pi/2$, we may without loss of generality assume that instead of a rectangle the billiard is a square and equal to $[0, \frac{1}{2}]^2$. The 'folding' map corresponding to this billiard is given by a two-variable function $f_2 : [0, \infty)^2 \to [0, \frac{1}{2}]^2$:

$$f_2(x, y) = (||x||, ||y||).$$

As we see, $f_2(x, y) = (f_1(x), f_1(y))$, which is why the billiard can be viewed as being a generalization of the setting of the Four Gap Theorem to two dimensions. The folding map f_2 applied to a line creates a billiard orbit. Let $\alpha > 0$, then

$$B_{[0,M]}^{\alpha} := \{(||x||, ||\alpha x||) : x \in [0, M]\}$$

describes a truncated billiard orbit that has initial slope α (the slope alternates between α and $-\alpha$). Let $\mathcal{A}_{[0,M]}^{\alpha}$ and $\mathcal{S}_{[0,M]}^{\alpha}$ denote the number of different areas respectively different shapes in the partition of $[0, \frac{1}{2}]^2$ induced by $B_{[0,M]}^{\alpha}$. Two shapes are different if one can not be obtained from the other by translating, rotating and reflecting. For orbits truncated in a boundary point, we will prove the following theorem:

Theorem 2. *Let $\alpha > 0$ and choose $M > 0$ such that $(||M||, ||\alpha M||) \in [0, \frac{1}{2}]^2 \setminus (0, \frac{1}{2})^2$. Then the billiard orbit $B^{\alpha}_{[0,M]}$ induces a partition of $[0, \frac{1}{2}]^2$ in polygons for which*

$$\mathcal{A}^{\alpha}_{[0,M]} \leq 13 \quad \textit{and} \quad \mathcal{S}^{\alpha}_{[0,M]} \leq 16.$$

These upper bounds are the best possible.

In this theorem the billiard is square, but the result for rectangular billiards easily follows since the square can be scaled to any rectangle without changing the ratios between the shapes. From now on, we will assume that M satisfies the condition in the theorem.

4. Orbit construction

We already have an explicit expression for the billiard orbit $B^{\alpha}_{[0,M]}$, but we will need a more tractable description. Therefore, in this section we present a rough intuitive outline of the way one can think of the geometry and construction of the billiard. The corresponding lemmata and their proofs are given in Section 6. Consider the unit square and draw a line starting from the lower left corner with slope α. The boundaries are now considered to be connected as in a torus, so when we reach it, the line continues at the opposite boundary. Equivalently, if one of the coordinates is about to exceed 1, we subtract 1. But this is exactly taking fractional parts in both coordinates. Therefore, after we have traversed the unit square N times, we have a set which can be expressed as

$$\{(\{x\}, \{\alpha x\}) : 0 \leq x < M\},$$

for some $M \in \mathbb{R}$. A plot of such a set is shown in the left panel of Figure 5.

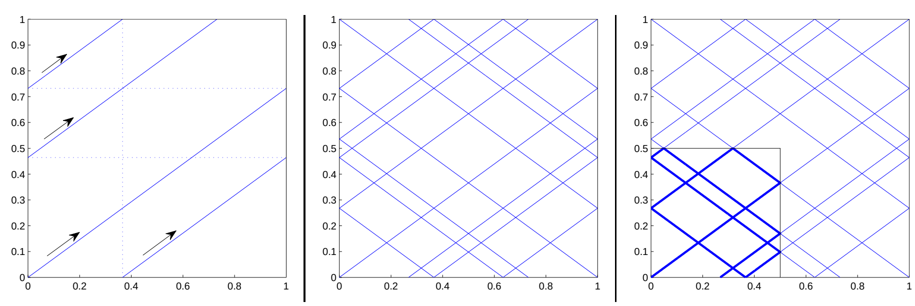

Figure 5: Construction of a billiard orbit in three steps. Here $N = 4$ and $\alpha = \sqrt{3} - 1$.

Now do the same starting from the other corners, traversing the square N times with a line either with slope α or $-\alpha$. Explicit expressions for these four sets (one for each corner) are given in Lemma 2. For an illustration, see the middle plot in Figure 5.

The key observation now is that intersection of all $4N$ lines with $[0, \frac{1}{2}]^2$ gives exactly a truncated billiard orbit with slope α, as is proved in Lemma 4. This fact is illustrated in the right plot in Figure 5. Obviously not all $4N$ lines actually contribute to the billiard orbit. However, there is a good reason to consider them all: the intercepts of the $2N$ lines with positive slope form a truncated orbit of a rotation on the interval $[-\alpha, 1]$, see Lemma 3. For the lines with negative slope a similar result holds. Having collected these insights, a simple counting argument suffices to obtain the upper bounds claimed in Theorem 2:

Proof of Theorem 2 Lemma 4 writes the billiard orbit as an intersection of the square $[0, \frac{1}{2}]^2$ with a set of lines. Let us concentrate on the lines with positive slope. By Lemma 3 the intercepts of these lines form a rotation orbit on the interval $[-\alpha, 1]$. So by Property 1 they induce a partition of this interval in subintervals of *at most* three different lengths. Denote the set of these lengths by $D := \{d_1, \ldots, d_n\}$, where $n \leq 3$. For the lines with negative slope, the intercepts are the numbers $1 - y_k$, $-N \leq k \leq N$. They induce a partition of $[0, 1 + \alpha]$ in subintervals having lengths in the same set D. It now follows that vertical distances between adjacent parallel lines are in the set D.

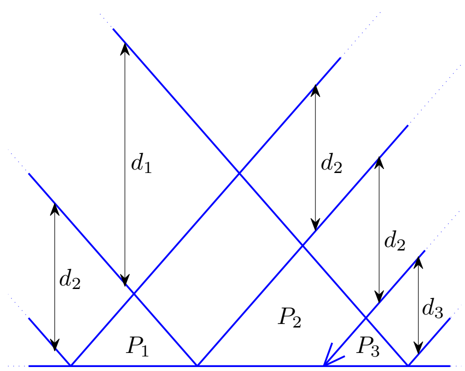

Figure 6: Local situation at the boundary where the orbit ends. Polygons of type 2 are triangular if the endpoint of the orbit is not one of the corners of the polygon (as is the case with P_1). There are only two shapes for which the endpoint of the orbit is one of the corners. One of them is still triangular (in this example P_3), the other is irregular (P_2).

We will distinguish between three types of polygons: those that have no side which is part of the boundary of $[0, \frac{1}{2}]^2$ (type 1), those that have exactly one such a side (type 2) and those that have two or more (type 3).

The polygons of type 1 must be parallelograms. The area of such a parallelogram is given by $d_i d_j / 2\alpha$ for some $d_i, d_j \in D$, and consequently they can have at most six different areas.

A polygon of type 2 that is triangular must be half of a rhombus of which the vertical diagonal has length $d \in D$, and therefore its area is $d^2/4a$. There is at

most one non-triangular type 2 polygon, as is explained in Figure 6. So polygons of type 2 can have at most four different areas.
Polygons of type 3 must be in one of the corners of $[0, \frac{1}{2}]^2$, but not in $(0,0)$ since the orbit starts there. So this gives at most three more areas.
Putting everything together, it turns out that the number of different areas is bounded by thirteen.
For the number of shapes a similar counting argument holds. The number of parallelogram shapes is again six, since reflections do not count. The triangles that are half of a rhombus can have at most six different shapes, since there are three types of rhombi which can be cut either horizontally or vertically. The rest of the argument doesn't change, so there are at most three more different shapes than different areas, which establishes the upper bound of at most sixteen different shapes.
The sharpness of these bounds follows from Example 1 in section 7. □

Remark 1. As the careful reader may have noted, the construction of the billiard orbit always gives a truncation on the left boundary or on the lower boundary of the square. So strictly speaking, Theorem 2 is not proved in full generality yet. Suppose we have an orbit truncated at the upper or right boundary. By removing the last linear part or adding the next linear part, we can transform this orbit into an orbit truncated at the left or lower boundary. This means that in the proof above, the rotation orbit on the interval $[-\alpha, 1]$ contains one element more or one less than the rotation orbit on $[0, 1+\alpha]$. Now Property 2 tells us that vertical distances between adjacent parallel lines can still have at most three different values, completing the proof.

5. Rational angles and a golden exception

Theorem 2 gives an upper bound for the number of different areas of shapes on the billiard table. Some natural questions remain. For example, what happens if α is rational? Can we prove sharper upper bounds under suitable conditions? In this section we explore these properties.
Obviously, taking α rational gives a special case. The first thing to note is that the orbit will be periodic: if $\alpha = p/q$, then for $x \in \mathbb{R}$

$$(||x+q||, ||\alpha(x+q)||) = (||x||, ||\alpha x||).$$

A bit less trivial is the following result.

Proposition 1. *The best upper bound for $\mathcal{A}^{\alpha}_{[0,M]}$ with $\alpha \in \mathbb{Q}$ is 13, but for all $\alpha \in \mathbb{Q}$ there is an M_0 such that $1 \leq \mathcal{A}^{\alpha}_{[0,M]} \leq 3$ for $M \geq M_0$. These bounds are sharp.*

Proof Note that the areas of the polygons continuously depend on α. So if we have an $\tilde{\alpha}$ and M such that $\mathcal{A}^{\tilde{\alpha}}_{[0,M]} = 13$, then we can find $\varepsilon > 0$ such that the upper bound of thirteen is reached for all $\alpha \in (\tilde{\alpha} - \varepsilon, \tilde{\alpha} + \varepsilon)$. Since this interval contains rationals, we see that rationality is not sufficient for a sharper upper

bound.
Since the orbit is periodic, the partition doesn't change anymore if M is large enough. Taking $\alpha = 1$ shows that 1 is a sharp lower bound for the limiting number of shapes. For the upper bound, suppose that $\alpha = p/q$. By Lemma 3 the intercepts satisfy

$$y_{p+q} = \left(1 + \frac{p}{q}\right)\left\{\frac{(p+q)p/q}{1+p/q}\right\} - \frac{p}{q} = -\frac{p}{q} = y_0.$$

It follows that the numbers y_k form a periodic rotation orbit on $[-\alpha, 1]$ and therefore the set D as defined in the proof of Theorem 2 contains only one length if M is large enough. If p and q are relative prime, then this length is $1/q$. Now a type 1 polygon is a rhombus with area $1/2pq$. Since there is no endpoint of the orbit anymore, a type 2 polygon is half of such a rhombus. Polygons in the corners are also triangular, because the orbit touches all sides of the square before becoming periodic. These triangles are quarters of the rhombus, thus having area $1/8pq$. This makes at most three different areas in total. To see that this upper bound is sharp, see Figure 7. □

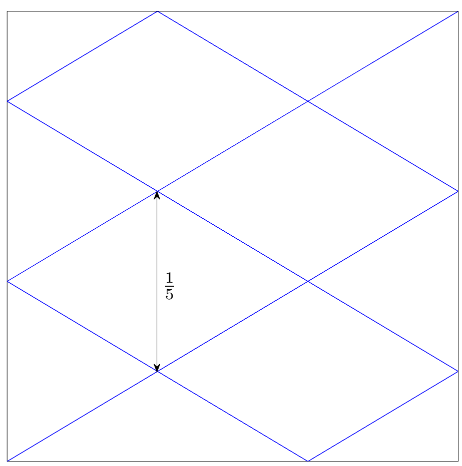

Figure 7: The periodic orbit for $\alpha = 3/5$. There are three different areas: the rhombi have area $1/(2 \cdot 3 \cdot 5) = 1/30$. The triangles have area $1/60$ or $1/120$.

Surprisingly, there exist irrational α for which the upper bound of thirteen different areas is never reached:

Proposition 2. *Let $\phi = (\sqrt{5}-1)/2$ denote the small golden mean. If $\alpha = \frac{1}{n+\phi}$ for some $n \in \mathbb{N}$, then $\mathcal{A}^{\alpha}_{[0,M]} \leq 12$.*

Proof Consider the numbers y_k that form a rotation orbit on $[-\alpha, 1]$. The partition of $[-\alpha, 1]$ induced by this orbit gives subintervals with lengths in a set D. This set D changes if we extend the orbit (i.e. we increase M): some lengths

will disappear and new lengths will be created. In [1] and [3] it was shown that the largest length is always the first to disappear. A new length only pops up if there are only two lengths in D, and the new length is the difference of these two existing lengths. Together with the fact that $1-\phi=\phi^2$, this is the basis of our argument.
Let $\alpha = 1/(n+\phi)$. From the way points are added to the rotation orbit it is clear that we can choose M such that $[-\alpha,1]$ will be partitioned in $n+1$ intervals of length α and an interval of length $1+\alpha-(n+1)\alpha=\phi\alpha$. This gives $D=\{\alpha,\phi\alpha\}$. Extending the orbit with one more point transforms D into $\left\{\alpha,\phi\alpha,\phi^2\alpha\right\}$ and this is the first time that D contains three lengths. Increasing M further, D will change into $\left\{\phi\alpha,\phi^2\alpha\right\}$ and then into $\left\{\phi\alpha,\phi^2\alpha,\phi^3\alpha\right\}$. An inductive argument suffices to show that the ratios between the lengths in D are preserved.
Recall that the areas of the parallelograms are determined by a product of two lengths in D. By the above reasoning, if $D=\{d_1,d_2,d_3\}$, then $d_1d_3=d_2^2$, which implies that the parallelograms can have at most five different areas. Consequently $\mathcal{A}^{\alpha}_{[0,M]}\leq 12$. □

6. Lemmata and their proofs

Let $\alpha>0$ be an irrational number and consider the halfline $l(x)=\alpha x, x\geq 0$. Let $S_1=[0,1)^2$ and define $S_2,S_3,S_4,\ldots$ to be the squares of the form $[k,k+1)\times[m,m+1)$, with k and m integers, that are consecutively traversed by the halfline, see Figure 8. Choosing an index N, there exists $M\in\mathbb{R}$ such that

$$\bigcup_{k=1}^{N} S_k\cap\{(x,\alpha x):x\geq 0\}=\{(x,\alpha x):0\leq x<M\}.$$

Taking fractional parts in both coordinates can be seen as mapping each of the squares S_k to $[0,1)^2$. Therefore, doing this for the above set gives

$$\{(\{x\},\{\alpha x\}):0\leq x<M\}=[0,1)^2\cap\bigcup_{k=1}^{N}\{(x,\alpha x+y_k):x\in\mathbb{R}\} \qquad (4)$$

for numbers y_k defined by the recursion

$$\begin{aligned} y_1 &= 0,\\ y_{k+1} &= \begin{cases} y_k+\alpha & \text{if } y_k<1-\alpha,\\ y_k-1 & \text{if } y_k>1-\alpha. \end{cases} \end{aligned} \qquad (5)$$

We will denote the set in (4) by A^{++}. The ++ superscript reflects the fact that we started with a halfline in the first quadrant, so both coordinates are positive. Doing similar operations to halflines in the second, third and fourth quadrant, we can define sets A^{-+}, A^{--} and A^{+-} respectively as follows:

$$\begin{aligned} A^{-+} &= \{(1-\{x\},\{\alpha x\}):0\leq x<M\},\\ A^{--} &= \{(1-\{x\},1-\{\alpha x\}):0\leq x<M\},\\ A^{+-} &= \{(\{x\},1-\{\alpha x\}):0\leq x<M\}. \end{aligned}$$

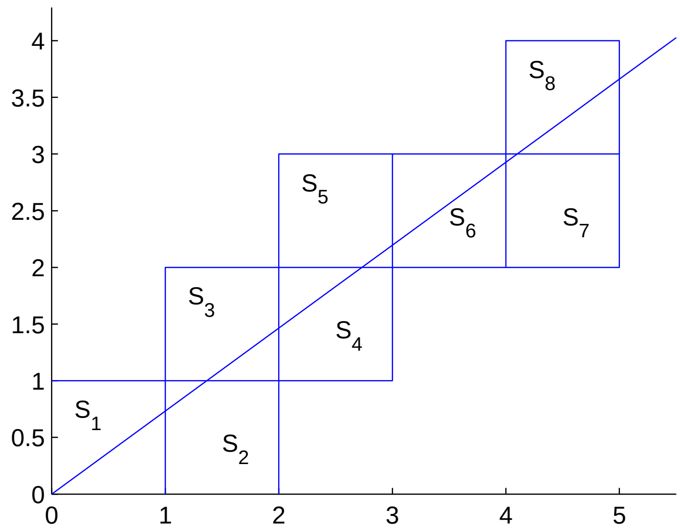

Figure 8: Construction of the squares S_k. Here $\alpha = \sqrt{3} - 1$, $N = 8$ and $M = 5$. The numbers y_k are approximately given by $y_0 = 0$, $y_1 \approx 0.732$, $y_2 \approx -0.268$, $y_3 \approx 0.464$,... Compare with Figure 5, left plot.

Taking the union of these four sets and intersecting with $[0, \frac{1}{2}]$ gives us a billiard orbit, as is proved in the lemma below.

Lemma 1. *The billiard orbit* $B^{\alpha}_{[0,M)}$ *satisfies*

$$B^{\alpha}_{[0,M)} = \bigcup_{u,v \in \{+,-\}} A^{uv} \cap [0, \frac{1}{2}]^2.$$

Proof Observe that

$$\begin{aligned} (||x||, ||\alpha x||) &= \Big(\min\Big\{\{x\}, 1 - \{x\}\Big\}, \min\Big\{\{\alpha x\}, 1 - \{\alpha x\}\Big\}\Big) \\ &= [0, \frac{1}{2}]^2 \cap \bigcup_{a \in \{\{x\}, 1-\{x\}\}} \bigcup_{b \in \{\{\alpha x\}, 1-\{\alpha x\}\}} (a, b), \end{aligned}$$

and now take the union over all $x \in [0, M)$. $\square$

In the next lemma expressions similar to (4) are derived for A^{-+}, A^{--} and A^{+-}.

Lemma 2. *Let* $y_{-k} = 1 - \alpha - y_k$ *for* $k = 1, 2, \ldots, N$. *Then*

$$A^{-+} = (0, 1] \times [0, 1) \cap \bigcup_{k=1}^{N} \{(x, -\alpha x + 1 - y_{-k}) : x \in \mathbb{R}\},$$

$$
\begin{aligned}
A^{--} &= (0,1]^2 \cap \bigcup_{k=1}^{N} \{(x, \alpha x + y_{-k}) : x \in \mathbb{R}\}, \\
A^{+-} &= [0,1) \times (0,1] \cap \bigcup_{k=1}^{N} \{(x, -\alpha x + 1 - y_k) : x \in \mathbb{R}\},
\end{aligned}
$$

Proof Define the functions $f, g, h : \mathbb{R}^2 \to \mathbb{R}^2$ by $f((x,y)) = (1-x, y)$, $g((x,y)) = (1-x, 1-y)$ and $h((x,y)) = (x, 1-y)$. Applying these functions to the left hand side of (4), we get $f(A^{++}) = A^{-+}$, $g(A^{++}) = A^{--}$ and $h(A^{++}) = A^{+-}$. On the other hand,

$$
\begin{aligned}
f(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}) &= \{(1-x, \alpha x + y_k) : x \in \mathbb{R}\} \\
&= \{(x, \alpha(1-x) + y_k) : x \in \mathbb{R}\} \\
&= \{(x, -\alpha x + 1 - y_{-k}) : x \in \mathbb{R}\},
\end{aligned}
$$

whence application of f to the right hand side of (4) leads to

$$
\begin{aligned}
f\Big([0,1)^2 \quad \cap \quad & \bigcup_{k=1}^{N} \{(x, \alpha x + y_k) : x \in \mathbb{R}\}\Big) \\
&= f\Big([0,1)^2\Big) \cap f\Big(\bigcup_{k=1}^{N} \{(x, \alpha x + y_k) : x \in \mathbb{R}\}\Big) \\
&= (0,1] \times [0,1) \cap \bigcup_{k=1}^{N} f\Big(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}\Big) \\
&= (0,1] \times [0,1) \cap \bigcup_{k=1}^{N} \{(x, -\alpha x + 1 - y_{-k}) : x \in \mathbb{R}\},
\end{aligned}
$$

so for A^{-+} we established the equality claimed in the lemma. The other two equalities for A^{--} and A^{+-} follow from a similar reasoning since

$$
\begin{aligned}
g(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}) &= \{(1-x, 1 - \alpha x - y_k) : x \in \mathbb{R}\} \\
&= \{(x, 1 - \alpha(1-x) - y_k) : x \in \mathbb{R}\} \\
&= \{(x, \alpha x + y_{-k}) : x \in \mathbb{R}\},
\end{aligned}
$$

and

$$
h(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}) = \{(x, 1 - \alpha x - y_k) : x \in \mathbb{R}\}.
$$

□

The numbers y_k and y_{-k} satisfy a nice relation, as is shown in the following lemma.

Lemma 3. *Let $y_0 = -\alpha$. Then the numbers $y_{-N}, \ldots, y_N$ form a rotation orbit on the interval $[-\alpha, 1]$. They are given by*

$$
y_k = (1+\alpha)\left\{\frac{k\alpha}{1+\alpha}\right\} - \alpha \quad \textit{for} \quad -N \leq k \leq N. \tag{6}
$$

Proof The recursion (5) can be rewritten as

$$y_{k+1} = (y_k + 2\alpha \bmod (1+\alpha)) - \alpha,$$

and therefore

$$\frac{y_{k+1}+\alpha}{1+\alpha} = \frac{y_k + 2\alpha}{1+\alpha} \bmod 1 = \left\{ \frac{y_k+2\alpha}{1+\alpha} \right\},$$

Letting $\tilde{y}_k = \frac{y_k+\alpha}{1+\alpha}, k = -N, \ldots, N$ and $\tilde{\alpha} = \frac{\alpha}{1+\alpha}$, this reduces to

$$\tilde{y}_{k+1} = \{\tilde{y}_k + \tilde{\alpha}\}.$$

Since $y_1 = 0$, we have $\tilde{y}_1 = \tilde{\alpha}$, which leads to

$$\tilde{y}_k = \{k\tilde{\alpha}\} \quad \text{for} \quad k \geq 1.$$

On the other hand, for $k \geq 1$,

$$\begin{aligned} \tilde{y}_{-k} &= \frac{y_{-k}+\alpha}{1+\alpha} = \frac{1-\alpha-y_k+\alpha}{1+\alpha} = \frac{1+\alpha}{1+\alpha} - \frac{y_k+\alpha}{1+\alpha} \\ &= 1 - \tilde{y}_k = 1 - \{k\tilde{\alpha}\} = \{-k\tilde{\alpha}\}, \end{aligned}$$

since $\tilde{\alpha}$ is irrational. By definition we have $\tilde{y}_0 = 0$, and hence

$$\tilde{y}_k = \{k\tilde{\alpha}\} \quad \text{for} \quad -N \leq k \leq N.$$

Solving for y_k gives the result. □

In Lemma 1 we already derived an expression for $B^\alpha_{[0,M)}$, but this is not so easy to analyze directly. In the next lemma we describe $B^\alpha_{[0,M]}$ as the union of two collections of lines intersected with $[0, \frac{1}{2}]^2$. All lines in the first collection have slope α and all lines in the second collection have slope $-\alpha$.

Lemma 4. *Let $l^+_k(x) = \alpha x + y_k$ and $l^-_k(x) = -\alpha x + 1 - y_k$. Then*

$$B^\alpha_{[0,M]} = [0, \frac{1}{2}]^2 \cap \bigcup_{u\in\{+,-\}} \bigcup_{k=-N}^{N} \{(x, l^u_k(x) : x \in \mathbb{R})\}$$

Proof This lemma will be proved by taking closures in the equation in Lemma 1.

$$\overline{B^\alpha_{[0,M)}} = B^\alpha_{[0,M]},$$

since $x \mapsto (||x||, ||\alpha x||)$ is a continuous function from $\mathbb{R}$ to $\mathbb{R}^2$. On the other hand,

$$\overline{[0, \frac{1}{2}]^2 \cap \bigcup_{u,v\in\{+,-\}} A^{uv}} = [0, \frac{1}{2}]^2 \cap \bigcup_{u,v\in\{+,-\}} \overline{A^{uv}},$$

and since A^{uv} is a finite collection of lines intersected by a 'half open' unit square its closure is the same collection of lines but now intersected by the closed square $[0,1]^2$. Therefore,

$$\overline{A^{++}} \cup \overline{A^{--}} = [0,1]^2 \cap \bigcup_{\substack{k=-N \\ k \neq 0}}^{N} \left\{(x, l_k^+(x)) : x \in \mathbb{R}\right\} \tag{7}$$

Now note that since $l_0^+(x) = \alpha x - \alpha$ we have

$$[0, \frac{1}{2}]^2 \cap \left\{(x, l_0^+(x)) : x \in \mathbb{R}\right\} = \emptyset.$$

Intersecting both sides of (7) with $[0, \frac{1}{2}]^2$ gives

$$[0, \frac{1}{2}]^2 \cap \overline{A^{++}} \cup \overline{A^{--}} = [0, \frac{1}{2}]^2 \cap \bigcup_{k=-N}^{N} \left\{(x, l_k^+(x)) : x \in \mathbb{R}\right\} \tag{8}$$

Analogously it follows that

$$[0, \frac{1}{2}]^2 \cap \overline{A^{+-}} \cup \overline{A^{-+}} = [0, \frac{1}{2}]^2 \cap \bigcup_{k=-N}^{N} \left\{(x, l_k^-(x)) : x \in \mathbb{R}\right\} \tag{9}$$

Combination of the last two equations gives the result. □

7. Sharpness of the bounds

In this section we present an example in which the upper bounds of Theorem 2 are reached. This proves sharpness of the bounds.

Example 1. Let $\alpha = \frac{\sqrt{10}}{7}$ and choose $N = 11$. The corresponding orbit is shown in Figure 9. Use Lemma 3 to find the numbers y_k and let

$$d_1 = y_{-2} - y_1 \approx 0.0965, \quad d_2 = y_{11} - y_{-2} \approx 0.0658, \quad d_3 = y_{-5} - y_{11} \approx 0.0307$$

denote the three different vertical distances between adjacent parallel lines. The areas of the shapes are of the following form:

$$\begin{array}{ll} \text{Shapes } I,\ II,\ III,\ IV,\ V,\ VI: & d_i d_j/2\alpha, \quad i \leq j \in \{1,2,3\} \\ \text{Shapes } VII,\ VIII,\ IX: & d_i^2/4\alpha, \quad i \in \{1,2,3\} \\ \text{Shape } X: & d_3 d_1/2\alpha - d_3^2/4\alpha \\ \text{Shapes } XI,\ XII,\ XIII: & d_i^2/8\alpha, \quad i \in \{1,2,3\} \end{array} \tag{10}$$

Calculating these thirteen areas indeed gives thirteen different values, where a precision of two decimals suffices. The flakes VII, $VIII$ and IX have the same areas as $VIIa$, $VIIIa$ and IXa respectively, so the maximal number of sixteen different shapes is also reached. We checked the calculations by using the outcomes to determine the area of $[0, \frac{1}{2}]^2$.

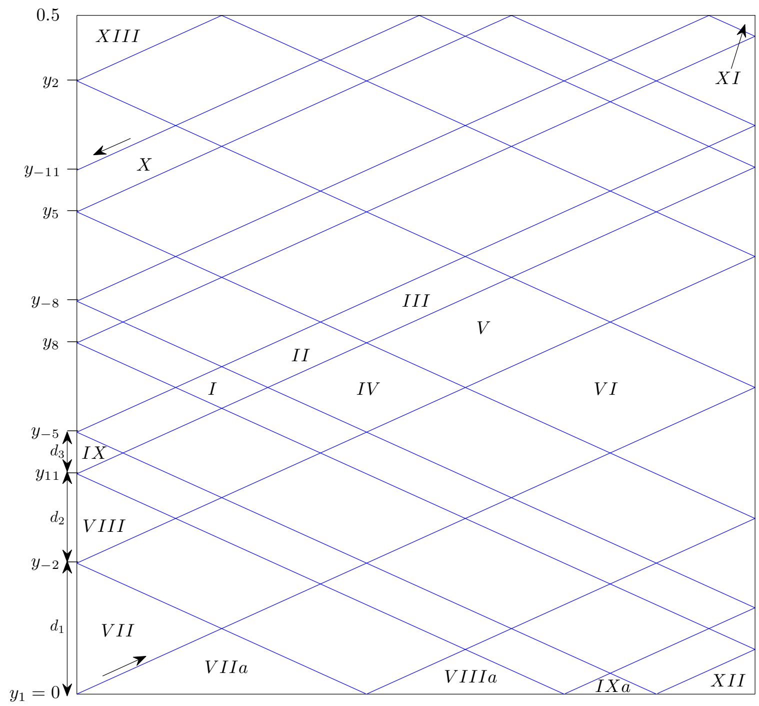

Figure 9: Thirteen different areas, sixteen different shapes.

Acknowledgement

The author thanks Michel Dekking and Cor Kraaikamp for their useful comments.

References

[1] Don, H. – *On the distribution of the distances of multiples of an irrational number to the nearest integer*, Acta Arith. 139 (2009), 253–264.

[2] Komornik, V. – Personal communication.

[3] van Ravenstein, T. – *The three gap theorem (Steinhaus conjecture)*, J. Austral. Math. Soc. Ser. A 45 (1988), no. 3, 360–370.

[4] Sós, V.T. – *On the distribution mod 1 of the sequence $n\alpha$*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 1 (1958), 127–134.