# UTD Sloan Math Pathways REU
# Geometry and Dynamics of Unusual Billiards

Niko Laohoo and Nick Robinson

July 2022

**Abstract**

Mathematical billiards has a long history with polygonal, circular, and elliptical tables. Beyond this, we can build unusual billiard tables as tables made up of confocal conics with reflex angles on the boundary. The motivation for this project originates from a problem posed by Dragović and Radnović in 2015. These unusual tables have interesting interactions with the segments connecting the arcs of conics leading to seemingly arbitrary situations where a cue ball admits periodic and non-periodic trajectories. For this paper, we focus on tables made up of concentric circles. For purely circular tables the relationship between rotation numbers and periodic trajectories is well-understood, however for unusual tables made up of several concentric circles little is known about how rotation numbers affect a cue ball's trajectory.

# 1 Introduction

Billiards is a very popular pastime enjoyed in many countries throughout the world. While physical billiards and mathematical billiards are both based on the same fundamental concept, some key differences exist.

1. In physical billiards, the cue ball has size and mass. In mathematical billiards, the cue ball is taken to be a point particle.
2. In physical billiards, friction and spin are important aspects of the game. In mathematical billiards, friction and spin are ignored.
3. In physical billiards, the cue ball has inelastic collisions with the boundary. In mathematical billiards, the cue ball is assumed to have elastic collisions with the boundary.

For mathematical billiards we define the billiards table as a 2-dimensional Riemannian sub-manifold of $\mathbb{R}^2$ with a piecewise-smooth boundary. When the cue ball collides with the boundary at some boundary point, its trajectory is reflected with respect to the tangent line at that boundary point such that the ball's incoming angle is the same as its outgoing angle.

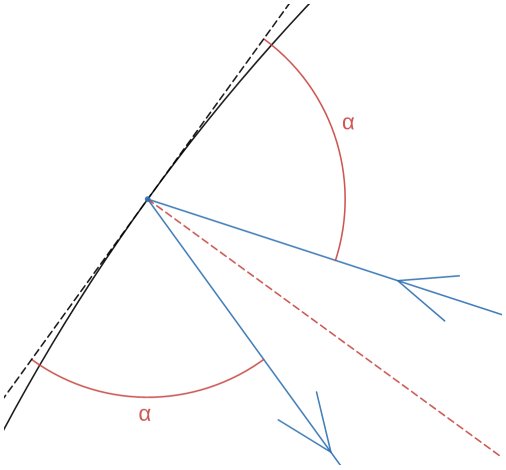



Figure 1: A visual of the billiard rule

Section 2 discusses the billiard trajectory rule for colliding with non-smooth points of the boundary. This project primarily explores tables made up of arcs of concentric circles. Given a billiard table we pick an initial point and angle which dictates the cue ball's trajectory. If the trajectory returns to its initial point after finitely many collisions with the boundary, then that trajectory is called periodic. Otherwise, the trajectory "densely" fills some region of the table.

## 1.1 $(t, \alpha)$ Coordinates

One possible approach to modeling trajectories in mathematical billiards consists of tracking the collision points along the boundary together with the outgoing angles of the trajectory at each of these collision points. To do this we use a manifold defined by Tabachnikov [Tab95]. Given a billiard table $M$ define the manifold $V$ to be the set of pairs $(x, v)$ where $x$ is a point on the boundary of $M$ and $v$ is a unit tangent vector at $x$ which points towards the interior of $M$. We can then model billiard trajectories in such a table by a function $T : V \to V$ which sends some boundary point and direction $(x, v)$ in $V$ to another boundary point and direction by moving the billiard particle from $x$ in the direction of $v$ until it collides with the boundary again, and then reflecting $v$ with respect to the tangent space at the new boundary point.

To perform computations we define the following coordinate system. Let $\gamma : [0, L) \to \partial M$ with parameter $t$ be a counter-clockwise, arc-length parametrization of the billiard table boundary ($L$ is the length of $\partial M$). The parameter $t$ provides the first coordinate for the coordinate system of $V$. Next, given a fixed $t$ we have the tangent vector $\gamma'(t) \in T_{\gamma(t)}(\partial M)$. Let $J$ be a 90° counter-clockwise rotation of a tangent vector in $\mathbb{R}^2$. We then define $\alpha$ to be the real number in $(0, \pi)$ such that $v = \cos(\alpha)\gamma'(t) + \sin(\alpha)J(\gamma'(t))$. Then $t$ and $\alpha$ define a coordinate system for $V$.

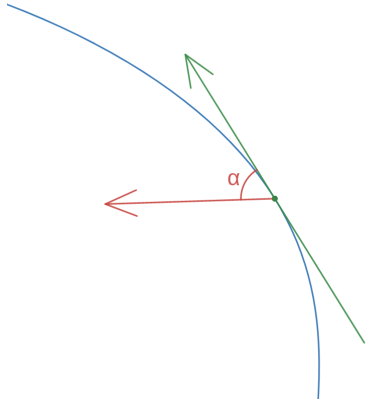



Figure 2: Visual of $(t, \alpha)$ coordinates

## 1.2 Circular Billiard Tables

For circular billiard tables we parametrize a circle of radius $R > 0$ by $C(t) = \left(R\cos\left(\frac{t}{R}\right), R\sin\left(\frac{t}{R}\right)\right)$. From this it can be shown that the billiard rule $T : V \to V$ in $(t, \alpha)$ coordinates is $T(t, \alpha) = (t+2\alpha R, \alpha)$. This shows that playing billiards in a circular billiard table is the same as rotating the initial boundary point about the circle by a fixed angle determined by the $\alpha$ coordinate of the initial direction. If $\alpha$ is a rational multiple of $\pi$ (i.e. there exists coprime integers $p, q$ such that $\alpha = \frac{p}{q}\pi$) then it can be shown that the $q$-fold composite of $T$ with itself is the identity on $V$. This means that $\pi$-rational choices of the initial angle produce orbits with period $q$.

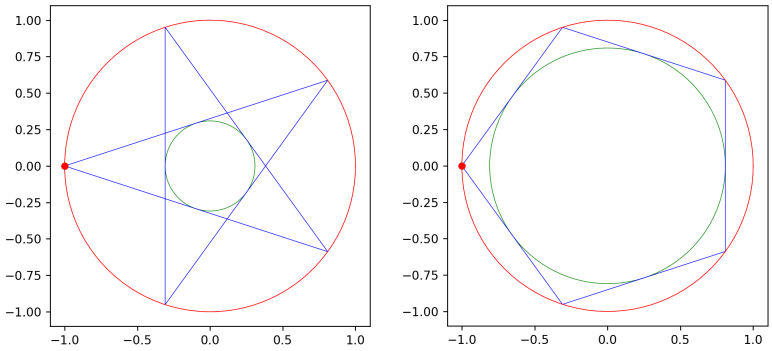



Figure 3: Circular tables with initial conditions $\alpha_0 = \frac{2}{5}\pi$ and $\alpha_0 = \frac{1}{5}\pi$ respectively

Of note in figure 3 are the inscribed circles that appear in green. These inscribed shapes are called *caustics* of the given billiard trajectory. A *caustic* is defined as a curve such that if a trajectory is tangent to it, then it remains tangent to it after every reflection [DR12]. Note that *caustics* are not guaranteed to exist for every billiard trajectory. Because of the nature of the billiard rule in circular

tables (specifically the fact that it is just a rotation), this caustic always appears for any choice of initial conditions with radius $R\cos(\alpha_0)$.

Even for $\pi$-irrational choices of $\alpha$, a caustic still arises. When the initial angle is not a rational multiple of $\pi$ then the region between the boundary and the caustic becomes "densely" filled by the trajectory. It looks as though the caustic is determined by the initial conditions, however the converse also holds. We can choose a caustic first (a circle concentric with the billiard boundary which has a smaller radius) and then study the set of billiard trajectories that have this fixed caustic.

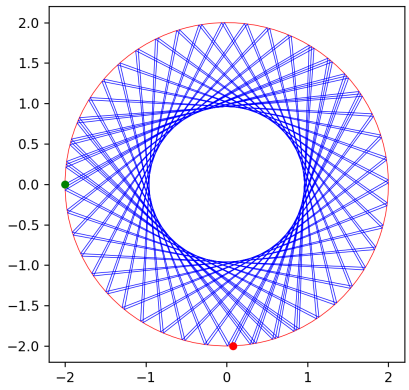



Figure 4: Billiard trajectory in circular table with $\pi$-irrational initial angle

## 1.3 Caustic-First Perspective

As stated before, we fix a circular billiard boundary and a concentric caustic and then study the set of billiard trajectories that produce the fixed caustic. To do this we first establish the idea of a rotation number. Let $r > 0$ be the radius of the fixed caustic and let $R > r$ be the radius of the billiard table boundary. We then produce figure 5.

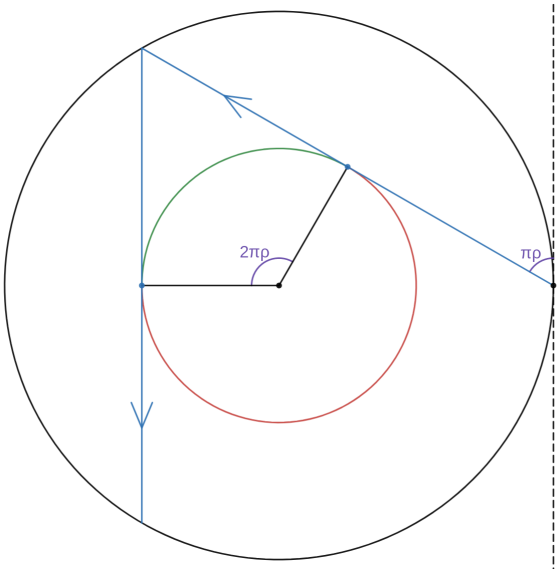



Figure 5: Rotation number visual

In figure 5 we choose concentric circles with radii $r = 1$ and $R = 2$, centered at the origin, with the initial boundary point being $(R, 0)$. With the fixed caustic we have two choices of initial direction. We choose the counter-clockwise direction. This partitions the caustic into the green and red regions seen in figure 5. The rotation number is the ratio of the length of the green region of the caustic to the circumference of the caustic. In this case the rotation number is $\rho = \frac{1}{3}$. In general for $R > r > 0$ we have that the rotation number is $\rho = \frac{1}{\pi} \arccos\left(\frac{r}{R}\right) \in \left(0, \frac{1}{2}\right)$.

If we have a rotation number $\rho \in \left(0, \frac{1}{2}\right)$ and a fixed caustic of radius $r > 0$, then we can play billiards by rotating the point tangent to the billiard trajectory on the caustic by $2\pi\rho$ radians. First, we parametrize the caustic $C_0$ by $C_0(\lambda) = (r\cos(2\pi\lambda), r\sin(2\pi\lambda))$. Given a point on $C_0$ the line through that point tangent to $C_0$ is a part of the billiard trajectory. To get the next part of the billiard trajectory we just add $\rho$ to $\lambda$. This results in a new billiard rule $g : C_0 \to C_0$ defined by $g(C_0(\lambda)) = C_0(\lambda + \rho)$ (if we wish to move clockwise we subtract $\rho$ instead of adding $\rho$). Thus playing billiards in a circular billiard table is equivalent to rotating a point on a fixed caustic by a fixed amount. This leads to the caustic-first approach for circular billiard tables. First we fix the radius $r > 0$ of the caustic and a rotation number $\rho \in \left(0, \frac{1}{2}\right)$. This gives us the radius of the billiard table as $R = \frac{r}{\cos(\pi\rho)}$. As we shall see, the caustic-first approach also works for other classes of billiard table.

### 1.4 Elliptical Billiard Tables

Instead of choosing concentric circles for the boundary and caustic, we can choose confocal conics for the boundary and the caustic. Given a choice of billiard table boundary and caustic, if we start a trajectory from the boundary of the billiard table which is tangent to the fixed caustic, then the trajectory will remain tangent to the fixed caustic after reflection. It should be noted that for ellipses, determining whether a choice of initial conditions produces a periodic trajectory is more complicated than in the case for concentric circles.

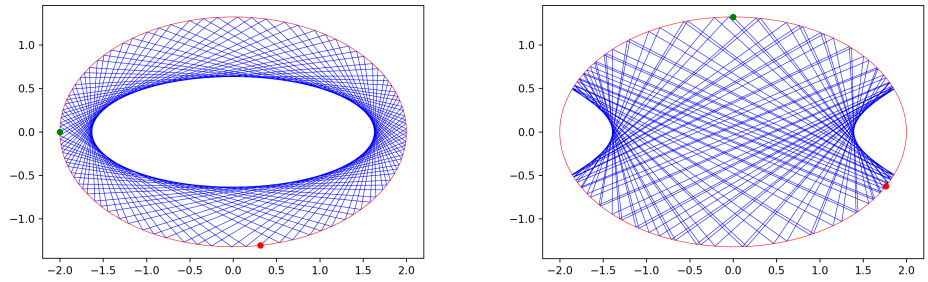



Figure 6: Two billiard trajectories in elliptical billiard tables with the caustics being confocal conics

## 2 Unusual Billiards

In this report, we explore problems posed by Dragović and Radnović [DR15]. We focus on tables that have boundaries made up of arcs of concentric circles. These tables are made up of two semicircles that have different radii connected by line segments. This leaves four points along the boundary that are not smooth: two reflex angles and two non-reflex angles. For the non-reflex angles, the reflected path can be defined to match the incoming path thanks to a limit that can be applied to the nearby trajectories [DR12]. For the reflex angles, no limit exists so if a trajectory hits a reflex corner the billiard trajectory stops; these trajectories are labeled “separatrices” [DR12].

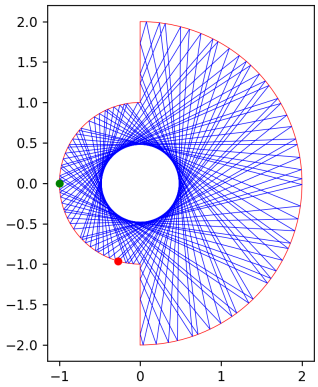


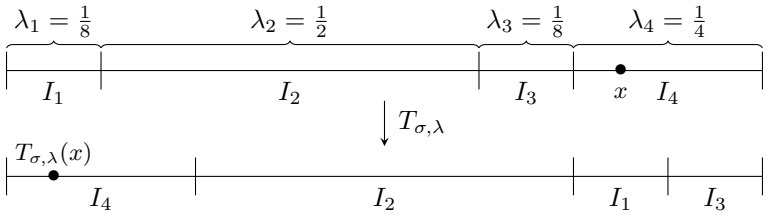

Figure 7: An example of a billiard trajectory in the unusual billiard table described in section 2. The circular caustic can be seen as the envelope of the billiard trajectory.

As with the case of circular billiard tables, it can be shown that any billiard trajectory in a billiard table of the described type produces a caustic and that this caustic is a concentric circle. To show this, first choose a circular caustic with radius smaller than the circles that make up the billiard table (as seen in figure 7). Given an oriented ray tangent to this caustic, it either collides with one of the semicircles or collides with a vertical wall. If the ray collides with a semicircle then the ray follows the usual reflection rule and its reflection is tangent to the caustic. If the ray collides with a vertical wall, it can be shown that the reflection of the ray produces a ray that is tangent to another point on the caustic.

Because of this we can use the caustic-first approach. This results in three parameters which completely describe a billiard table of this type. First is the radius of the caustic $r > 0$ which is usually fixed to 1 without loss of generality, after a possible rescaling. The next 2 parameters are the rotation numbers of the semicircles which make up the billiard table. These are denoted $\rho_1, \rho_2 \in \left(0, \frac{1}{2}\right)$ with $\rho_1 < \rho_2$. $\rho_1$ determines the radius of the smaller semicircle on the left $R_1 = \frac{r}{\cos(\pi\rho_1)}$ and $\rho_2$ determines the radius of the larger semicircle on the right $R_2 = \frac{r}{\cos(\pi\rho_2)}$. From this information a billiard rule can be made that sends points on the caustic to other points on the caustic such that the tangent lines of the points on the caustic produced by this billiard rule correspond to a billiard trajectory. The caustic is labeled $C_0$, the smaller semicircle on the left is labeled $C_{\rho_1}$, and the larger semicircle on the right is labeled $C_{\rho_2}$.

## 2.1 Interval Exchange Transformations

An interval exchange transformation is a map from some bounded interval of $\mathbb{R}$ to itself that permutes pre-defined subintervals which partition this bounded interval. For natural number $n \geq 1$ and an interval $[a, b)$ with $a < b$, we define a vector $\lambda = (\lambda_1, \ldots, \lambda_n)$ (with $\lambda_i > 0$ for $i = 1, \ldots, n$) and a permutation $\sigma$ on $\{1, \ldots, n\}$ such that $\sum_{i=1}^{n} \lambda_i = b - a$. From this we define $a_i = a + \sum_{1 \leq j < i} \lambda_j$ and subintervals $I_1, \ldots, I_n$ of $[a, b)$ defined by $I_i = [a_i, a_i + \lambda_i)$. We then define a map $T_{\sigma,\lambda} : [a, b) \to [a, b)$ which permutes the subintervals $I_1, \ldots, I_n$ in $[a, b)$ such that the $i$th subinterval from the left endpoint of $[a, b)$ becomes the $\sigma(i)$th subinterval from the left endpoint of $[a, b)$. As an example, let $n = 4$, $[a, b) = [0, 1)$, $\lambda = \left(\frac{1}{8}, \frac{1}{2}, \frac{1}{8}, \frac{1}{4}\right)$ and $\sigma = (1\ 3\ 4)$.




In the above picture we see that $x = 0.8125$ (which is in subinterval $I_4$) is mapped to 0.0625 by $T_{\sigma,\lambda}$. We can clearly see that $T_{\sigma,\lambda}$ has the effect of permuting the subintervals $I_1, \ldots, I_4$. Of note is that the distance of $x$ to the left and right endpoints of $I_4$ is the same as the distance of $T_{\sigma,\lambda}(x)$ from the left and right endpoints of $T_{\sigma,\lambda}(I_4)$.

The simplest non-trivial case is when $n = 2$ and $\sigma = (1\ 2)$. The vector $\lambda$ can be uniquely determined by a number $\nu \in (0, 1)$ with $\lambda = (\nu, 1 - \nu)$. Let $\lambda = \left(\frac{1}{4}, \frac{3}{4}\right)$ and $\sigma = (1\ 2)$. Let $\tilde{\lambda} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{2}\right)$ and $\tilde{\sigma} = (1\ 2\ 3)$. If we then express $T_{\sigma,\lambda}$ and $T_{\tilde{\sigma},\tilde{\lambda}}$ on $[0, 1)$ we get the following

$$T_{\sigma,\lambda}(x) = \begin{cases} x + \frac{3}{4} \text{ if } x \in \left[0, \frac{1}{4}\right) \\ x - \frac{1}{4} \text{ if } x \in \left[\frac{1}{4}, 1\right) \end{cases} \tag{2.1}$$

$$T_{\tilde{\sigma},\tilde{\lambda}}(x) = \begin{cases} x + \frac{3}{4} \text{ if } x \in \left[0, \frac{1}{4}\right) \\ x - \frac{1}{4} \text{ if } x \in \left[\frac{1}{4}, 1\right) \end{cases} \tag{2.2}$$

Clearly the pairs $(\sigma, \lambda)$ and $(\tilde{\sigma}, \tilde{\lambda})$ produce the same map, yet the pair $(\tilde{\sigma}, \tilde{\lambda})$ describes one more interval than the pair $(\sigma, \lambda)$. This leads to the following degeneracy condition for the permutation.

**Definition 2.1.** *Let $\sigma$ be a permutation on $\{1, \ldots, n\}$. $\sigma$ is said to be degenerate if there exists a $k = 1, \ldots, n - 1$ such that $\sigma(k) + 1 = \sigma(k + 1)$.*

By only choosing non-degenerate permutations we can avoid situations where two distinct $(\sigma, \lambda)$ pairs produce the same interval exchange transformation.

## 2.2 Ergodic Maps

In this section we introduce definitions related to ergodic theory. The ergodicity of a measure invariant map can be used to find where a map has periodic orbits for certain choices of initial point.

**Definition 2.2.** *Let $X$ be a set and $\mathcal{B}$ a $\sigma$-algebra on $X$. Then the pair $(X, \mathcal{B})$ is a* measurable space. *A* measure *on $(X, \mathcal{B})$ is a function $\mu$ from $\mathcal{B}$ to $[0, \infty)$ or $[0, \infty]$ such that*

1. *for all $E \in \mathcal{B}$, $\mu(E) \geq 0$,*
2. *$\mu(\varnothing) = 0$, and*
3. *if $\{E_k\}_{k=1}^{\infty}$ is a countably infinite collection of elements in $\mathcal{B}$ then $\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$.*

*The triple $(X, \mathcal{B}, \mu)$ is called a* measure space.

The following definitions are modified from Walters to extend to measure spaces which are not probability spaces [Wat82].

**Definition 2.3.** *Given a measurable space $(X, \mathcal{B})$, a map $T : X \to X$ is* measurable *if for all $A \in \mathcal{B}$ then $T^{-1}(A) \in \mathcal{B}$. Furthermore, given a measure $\mu$ on $(X, \mathcal{B})$, $T$ is* $\mu$-invariant *if for all $A \in \mathcal{B}$ then $\mu(T^{-1}(A)) = \mu(A)$.*

**Definition 2.4.** *Let $(X, \mathcal{B})$ be a measurable space. Let $\mu : \mathcal{B} \to [0, \infty)$ be a measure on $(X, \mathcal{B})$ with $\mu(X)$ being non-zero and finite and let $T : X \to X$ be a $\mu$-invariant map. $T$ is called $\mu$-ergodic when for all $A \in \mathcal{B}$ we have that $T^{-1}(A) = A$ implies either $A$ has measure 0 or full measure.*

As an example, consider the measure space $(X, \mathcal{B}, m)$ with $X = \{1, \ldots, n\}$ $(n \geq 2)$, $\mathcal{B} = \mathcal{P}(X)$, and $m$ the counting measure. Note that because the chosen $\sigma$-algebra is the power set of $X$, every map from $X$ to $X$ is measurable.

**Lemma 2.1.** *A map $T : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is $m$-invariant if and only if $T$ is bijective.*

*Proof.* Suppose $T$ is $m$-invariant. Let $b \in \{1, \ldots, n\}$. By hypothesis we have that $m(T^{-1}(\{b\})) = m(\{b\}) = 1$. This means there is some $a \in X$ such that $T^{-1}(\{b\}) = \{a\}$ since the set $T^{-1}(\{b\})$ contains exactly 1 element. This means there is exactly 1 element of $X$ such that $T(a) = b$. This shows that $T$ is bijective.

Suppose $T$ is bijective. Then there exists an inverse map $T^{-1}$ which is also a bijection. Let $A \in \mathcal{B}$. The cardinality of $A$ and the image of $A$ through bijection $T^{-1}$ have the same cardinality, so $T$ is $m$-invariant. □

**Theorem 2.2.** *$T : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is $m$-ergodic if and only if $T$ has one orbit (i.e. for every $x, y \in \{1, \ldots, n\}$ there exists a natural number $k \in \mathbb{N}$ such that $T^k(x) = y$).*

*Proof.* Suppose that $T$ is $m$-ergodic. This means $T$ is $m$-invariant and for every $A \in \mathcal{B}$ if $T(A) = A$ then $A$ is either empty $(m(A) = 0)$ or $A = X$ $(m(A) = m(X))$. Note that because of lemma 2.1, $T$ being $m$-invariant implies $T$ is bijective. This means for $A \in \mathcal{B}$, the statements $T(A) = A$ and $T^{-1}(A) = A$ are equivalent. Since $T$ is a bijection on $\{1, \ldots, n\}$ it is a permutation in $S_n$. Define the subgroup $H_T$ of $S_n$ by $H_T = \{T^k \mid k \in \mathbb{Z}\}$. Consider how $H_T$ acts on $\{1, \ldots, n\}$ by permutation. For $k \in \{1, \ldots, n\}$ we have that an orbit is as follows

$$\mathcal{O}_k = \{T^l(k) \in \{1, \ldots, n\} \mid l \in \mathbb{Z}\}$$

Obviously $T(\mathcal{O}_k) = \mathcal{O}_k$, so either $\mathcal{O}_k$ is empty or $\mathcal{O}_k$ is $\{1, \ldots, n\}$. Since $\mathcal{O}_k$ is not empty we have that $\mathcal{O}_k = \{1, \ldots, n\}$ for every $k \in \{1, \ldots, n\}$. This shows that $H_T$ acting on $\{1, \ldots, n\}$ by permutation has one orbit which is equivalent to $T$ having one orbit.

Suppose that $T$ has one orbit. Let $A \in \mathcal{B}$ such that $A$ is non-empty and $A$ is a proper subset of $X$. Suppose $T(A) = A$. This means $T^k(A) = A$ for every $k \in \mathbb{N}$. Because $A$ is non-empty and $X \setminus A$ is non-empty there exists elements $x \in A$ and $y \notin X \setminus A$. Because $T$ has one orbit there exists a natural number $l \in \mathbb{N}$ such that $T^l(x) = y$. This is sufficient to show that $T^l(A) \neq A$ which contradicts the statement that $T^k(A) = A$ for every $k \in \mathbb{N}$. Thus our assumption that $T(A) = A$ is false, meaning $T(A) \neq A$. This shows that $T$ is $m$-ergodic. □

In this report, we will express the problem of unusual billiards in terms of deciding whether a given interval exchange transformation is ergodic with respect to the Lebesgue measure. Lastly, we write theorem 1.14 from Walters called the Birkhoff Ergodic Theorem [Wat82].

**Definition 2.5.** *Let $(X, \mathcal{B})$ be a measurable space, $\mu$ a measure on $(X, \mathcal{B})$ such that $\mu(X)$ is non-zero and finite, and let $T : X \to X$ be a $\mu$-invariant map. Let $f : X \to \mathbb{R}$ be a $\mu$-integrable function. We define the time mean of $f$ with respect to $T$ at $x \in X$ as*

$$\hat{f}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$$

*where defined and define the space mean of $f$ as*

$$\bar{f} = \frac{1}{\mu(X)} \int_X f d\mu$$

**Theorem 2.3** (Birkhoff Ergodic Theorem)**.** *Let $T : X \to X$ be a $\mu$-invariant map and $f : X \to \mathbb{R}$ a $\mu$-integrable function. Then $\hat{f}(x)$ converges pointwise, $\hat{f} \circ T = \hat{f}$, and if $\mu(X)$ is finite then $\int_X \hat{f} d\mu = \int_X f d\mu$.*

**Remark.** *From a remark by Walters about the Birkhoff Ergodic Theorem, we can state the following: if $T$ is $\mu$-ergodic and $\mu(X)$ finite and non-zero then for any $\mu$-integrable function $f : X \to \mathbb{R}$, the set $\left\{x \in X \mid \hat{f}(x) = \bar{f}\right\}$ has full measure.*

The contrapositive of the above remark provides an applicable way to show that a given map $T : X \to X$ is not ergodic. Combined with the next theorem we can then apply both to billiards.

**Theorem 2.4.** *Let $\mu$ be the Lebesgue measure on $[a, b)$ and $T : [a, b) \to [a, b)$ a $\mu$-invariant map. Define $q = \inf\{n \in \mathbb{N} \mid T^n(x) = x \text{ for every } x \in [a, b)\}$ (i.e. $q$ is the smallest natural number such that $T^q$ is identical to the identity on $[a, b)$). If $q$ is finite then $T$ is not $\mu$-ergodic.*

*Proof.* Let $n \in \mathbb{N}$, $f : [a, b) \to \mathbb{R}$ be some $\mu$-integrable function, and $r$ the unique integer such that $n \equiv r \mod q$ and $0 \leq r < q$. Notice for $x \in [a, b)$

$$\begin{aligned}\sum_{k=0}^{n-1} f(T^k(x)) &= \underbrace{f(x) + f(T(x)) + \cdots + f(T^{q-1}(x))}_{\frac{n-r}{q} \text{ times}} + \cdots + f(x) + f(T(x)) + \cdots + f(T^{r-1}(x)) \\ &= \frac{n-r}{q} \sum_{k=0}^{q-1} f(T^k(x)) + \sum_{k=0}^{r-1} f(T^k(x))\end{aligned}$$

Therefore

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \left(1 - \frac{r}{n}\right) \frac{1}{q} \sum_{k=0}^{q-1} f(T^k(x)) + \frac{1}{n} \sum_{k=0}^{r-1} f(T^k(x))$$

When we take the limit as $n$ increases unbounded, the $\sum_{k=0}^{r-1} f(T^k(x))$ term is finite so $\frac{1}{n} \sum_{k=0}^{r-1} f(T^k(x))$ vanishes. Because $r$ is bounded by $q$ which is fixed, $\frac{r}{n}$ vanishes as $n$ increases unbounded. With this we have the following

$$\hat{f}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \frac{1}{q} \sum_{k=0}^{q-1} f(T^k(x))$$

Recall that $\bar{f} = \frac{1}{\mu(X)} \int_X f d\mu = \frac{1}{b-a} \int_X f d\mu$. Let $s$ be a real number such that $a < s < b$ and choose $f = \chi_{[a,s)}$ (that is, choose $f$ to be the characteristic function of $[a, s)$). This means $\bar{f} = \frac{s-a}{b-a}$. For a fixed $x \in [a, b)$ the value of $q\hat{f}(x)$ is the number of times a single orbit determined by $T$ and starting at $x$ visits the interval $[a, s)$. This means for a fixed choice of $x \in [a, b)$ there is an integer $l = 0, \ldots, q$ such that $\hat{f}(x) = \frac{l}{q}$. Choose $s = \frac{b-a}{2q} + a$. Then $\bar{f} = \frac{1}{2q}$. There does not exist an integer $l = 0, \ldots, q$ such that $\frac{l}{q} = \frac{1}{2q}$ so $\hat{f}$ is never equal to $\bar{f}$. Therefore we have found a $\mu$-integrable function such that the set of values in $[a, b)$ where $\hat{f}$ is equal to $\bar{f}$ is less than full measure. By the contrapositive of a remark about theorem 2.4 this means that $T$ is not $\mu$-ergodic. □

In the next section we will express caustic-first billiards in the language of interval exchange transformations. By doing this we can then apply ergodic theory to find results about the behavior billiard trajectories in unusual billiards.

## 2.3 Modelling Billiards with Interval Exchange Transformations

### 2.3.1 Circular Billiards

To give an example of how billiards can be modeled using interval exchange transformations, we express producing a billiard trajectory in a circular billiard table as an interval exchange transformation. Let $\rho \in \left(0, \frac{1}{2}\right)$, $\lambda = (\rho, 1-\rho)$, and $\sigma = (1\ 2)$. The interval exchange transformation associated with $(\sigma, \lambda)$ on $[0, 1)$ is written below.

$$T_{(\sigma,\lambda)}(x) = \begin{cases} x + 1 - \rho & x \in [0, \rho) \\ x - \rho & x \in [\rho, 1) \end{cases}$$

Let $C_0$ be a circular caustic of radius 1 centered at the origin. Parametrize $C_0$ by

$$C_0(t) = (\cos(2\pi t), \sin(2\pi t))$$

Notice this parametrization of $C_0$ has a period of 1 (i.e. $C_0(t+1) = C_0(t)$ for every $t \in \mathbb{R}$). Using this we observe the following:

1. For every $t \in [0, \rho)$, $C_0(T(t)) = C_0(t - \rho)$.
2. For every $t \in [\rho, 1)$, $C_0(T(t)) = C_0(t - \rho)$.

This shows that for every $t \in [0, 1)$ that $C_0(T(t)) = C_0(t - \rho)$. Furthermore for $k \in \mathbb{N}$ we have that $C_0(T^k(t)) = C_0(t - k\rho)$. This exactly corresponds to the billiard rule $g : C_0 \to C_0$ described in the caustic-first approach to circular billiards in section 1.3. This means proving statements about the periodicity of $T_{(\sigma,\lambda)}$ will give results about the periodicity of billiard trajectories in circular billiards.

### 2.3.2 Unusual Billiards

To describe unusual billiards using interval exchange transformations requires a more nuanced approach. Due to the piecewise nature of the boundary, both the direction and location of the of the billiard trajectory at any given time have to be considered more carefully. Recall from section 2 that an unusual billiard table is determined by a choice of rotation numbers $\rho_1, \rho_2 \in \left(0, \frac{1}{2}\right)$ with $\rho_1 < \rho_2$ and a caustic with radius fixed to 1. Given these parameters, there are four distinct events for a billiard trajectory to encounter at any given moment:

1. A trajectory collides with a boundary point on $C_{\rho_1}$.
2. A trajectory first collides with a boundary point on $C_{\rho_2}$ and then becomes tangent to the caustic again before the next collision.
3. A trajectory first collides with one of the vertical walls then collides a boundary point on $C_{\rho_2}$.
4. A trajectory first collides a boundary point on $C_{\rho_2}$ then collides with one of the walls.

By choosing a point on the caustic, we produce a line which is tangent to this point on the caustic. By orienting the line (giving it a direction) we create an oriented ray tangent to the specified point of the caustic. This choice of oriented ray determines the event encountered by the trajectory as described above. To formalize this, we first find for which points and directions of the caustic each of these events occur.

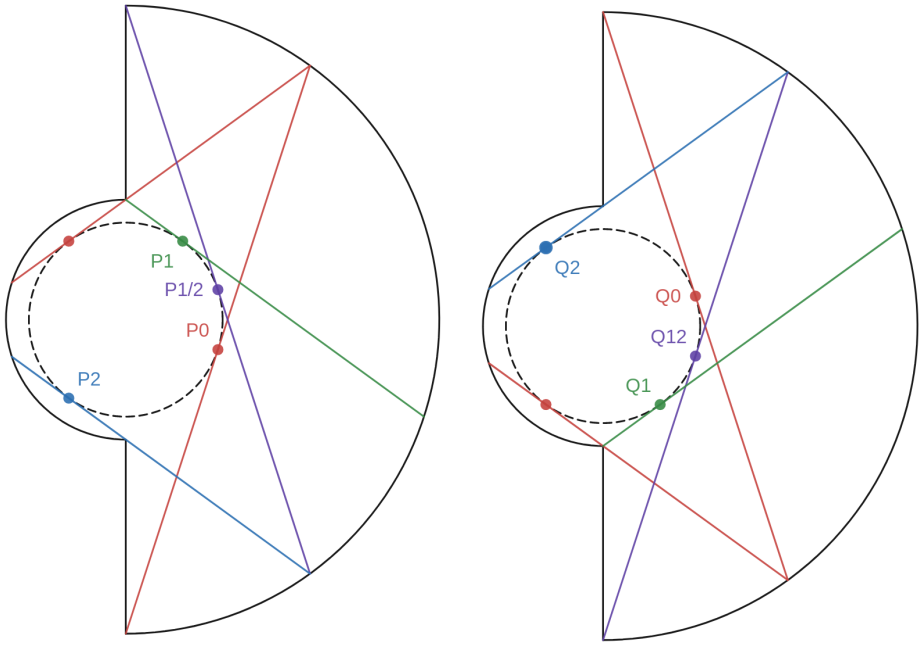



Figure 8: Visual of important points on a caustic for unusual billiards ($\rho_1 = \frac{1}{5}$ and $\rho_2 = \frac{2}{5}$).

To formalize this we use an approach used by Dragović [DR12]. Firstly, define the value $v = \frac{1}{4} + \frac{\rho_1}{2} - \rho_2$. On the boundary there are four critical points: the two non-reflex corners and the two reflex corners. Consider all trajectories tangent to the caustic $C_0$ moving counter-clockwise. By finding which points on the caustic produce counter-clockwise trajectories which collide with the critical boundary points, we can produce regions in which we observe each of the four events. We show this in the left of figure 8.

1. Given a counter-clockwise trajectory tangent to a point between $P_1$ and $P_2$ (counter-clockwise), the trajectory hits $C_{\rho_1}$.
2. Given a counter-clockwise trajectory tangent to a point between $P_2$ and $P_0$, the trajectory hits $C_{\rho_2}$. After reflection, the trajectory will become tangent to the caustic again before the next collision.
3. Given a counter-clockwise trajectory tangent to a point between $P_0$ and $P_{\frac{1}{2}}$, the trajectory collides with $C_{\rho_2}$ first then collides with the upper vertical wall.
4. Given a counter-clockwise trajectory tangent to a point between $P_{\frac{1}{2}}$ and $P_1$, the trajectory collides with the upper vertical wall then collides with $C_{\rho_2}$.

Note that if a counter-clockwise trajectory is tangent to $P_{\frac{1}{2}}$ then the direction of the trajectory after reflection is simply reversed. We find the following:

$$
\begin{aligned}
P_0 &= C_0(v) \\
P_{\frac{1}{2}} &= C_0\left(v + \frac{1}{2}(\rho_2 - \rho_1)\right) \\
P_1 &= C_0(v + \rho_2 - \rho_1) \\
P_2 &= C_0\left(v + \rho_2 - \rho_1 + \frac{1}{2}\right)
\end{aligned}
$$

A similar process can be applied to clockwise trajectories using the points $Q_0$, $Q_{\frac{1}{2}}$, $Q_1$, and $Q_2$. With this we define functions $p : C_0 \to [0,1)$ and $q : C_0 \to [-1,0)$ such that we have the following

$$\begin{aligned} p(P_0) &= 0, & q(Q_0) &= -1, \\ p\left(P_{\frac{1}{2}}\right) &= \frac{1}{2}(\rho_2 - \rho_1), & q\left(Q_{\frac{1}{2}}\right) &= -1 + \frac{1}{2}(\rho_2 - \rho_1), \\ p(P_1) &= \rho_2 - \rho_1, & q(Q_1) &= -1 + \rho_2 - \rho_1, \\ p(P_2) &= \rho_2 - \rho_1 + \frac{1}{2}, & q(Q_2) &= -1 + \rho_2 - \rho_1 + \frac{1}{2}. \end{aligned}$$

Furthermore, we have functions $p^{-1} : [0,1) \to C_0$ and $q^{-1} : [-1,0) \to C_0$ defined by $p^{-1}(t) = C_0(t+v)$ for $t \in [0,1)$ and $q^{-1}(t) = C_0(-(t+v))$ for $t \in [-1,0)$. This gives a coordinate system of oriented rays which are tangent to the fixed caustic $C_0$. Each oriented ray is associated with a unique real number in $[-1,1)$ through means of using functions $p$ and $q$. Specifically, if $t \in [0,1)$ then the associated ray is described by $((x,y) - p(t)) \cdot p(t) = 0$ moving in the counter-clockwise direction and if $t \in [-1,0)$ then the associated ray is described by $((x,y) - q(t)) \cdot q(t) = 0$ moving in the clockwise direction.

With all the tools above, we were able to find a general billiard rule $g : C_0 \to C_0$ defined almost everywhere on $C_0$ which works for any sign of $v$. First define $\bmod : \mathbb{R} \times \mathbb{R}^+ \to [0,m)$ by "$x \mod m = y$" if and only if there exists an integer $k \in \mathbb{Z}$ such that $x + km = y$ and $y \in [0,m)$. Define $j : \mathbb{R} \to \mathbb{R}$ by $j(x) = \frac{1}{\pi}\arccos(|\sin(2\pi x)|)$. We can then write how $g$ maps $t \in [-1,1)$.

$$g(t) = \begin{cases} [[-(t+\rho_2) - 2v + j(t+\rho_2+v)]] \mod 1] - 1 & 0 < t < \frac{1}{2}(\rho_2 - \rho_1) \\ [[-(t+2v)] \mod 1] - 1 & t = \frac{1}{2}(\rho_2 - \rho_1) \\ \left[\left[-\frac{1}{4} - v - \frac{1}{2}j(t+v) + \rho_2\right] \mod 1\right] - 1 & \frac{1}{2}(\rho_2 - \rho_1) < t < \rho_2 - \rho_1 \\ [t + \rho_1] \mod 1 & \rho_2 - \rho_1 < t < \rho_2 - \rho_1 + \frac{1}{2} \\ [t + \rho_2] \mod 1 & \rho_2 - \rho_1 + \frac{1}{2} < t < 1 \\ [-(t+\rho_2) - 2v + j(t+\rho_2+v)] \mod 1 & -1 < t < -1 + \frac{1}{2}(\rho_2 - \rho_1) \\ [-(t+2v)] \mod 1 & t = -1 + \frac{1}{2}(\rho_2 - \rho_1) \\ \left[-\frac{1}{4} - v - \frac{1}{2}j(t+v) + \rho_2\right] \mod 1 & -1 + \frac{1}{2}(\rho_2 - \rho_1) < t < -1 + \rho_2 - \rho_1 \\ [[t + \rho_1] \mod 1] - 1 & -1 + \rho_2 - \rho_1 < t < -1 + \rho_2 - \rho_1 + \frac{1}{2} \\ [[t + \rho_2] \mod 1] - 1 & -1 + \rho_2 - \rho_1 + \frac{1}{2} < t < 0 \end{cases} \tag{2.3}$$

Note $g$ is not defined for $t = 0, \rho_2 - \rho_1, \rho_2 - \rho_1 + \frac{1}{2}$ (associated with counter-clockwise rays) and $t = -1, -1 + \rho_2 - \rho_1, -1 + \rho_2 - \rho_1 + \frac{1}{2}$ (associated with clockwise ray).

It was shown by Dragović and Radnović that an interval exchange transformation can be produced that exactly corresponds to unusual billiards and the above coordinate system [DR12]. There are three distinct interval exchange transformations depending on if $v > 0$, $v < 0$, or $v = 0$. By finding results about interval exchange transformations we can find results about billiard trajectories in unusual billiards.

### 2.4 Ergodicity of Interval Exchange Transformations

As stated before, finding results about the ergodicity of interval exchange transformations will give results about the billiard trajectories in unusual billiards. The case of interval exchange transformations for $n = 2$ is well understood, since this has been shown to be the same as circular billiards. Future research will focus on when interval exchange transformations are ergodic for $n > 2$.

## 3 Simulations and Statistical Analysis

A portion of this project was dedicated to creating proper simulations to run experiments on. Two different simulations were created in Python, one running billiards experiments geometrically and one algebraically.

For the geometric program, a scene is defined which describes the boundary of a billiard table in the plane. An initial point and direction in the plane is chosen. Once these are fixed, the program uses ray marching to simulate a billiard ball traveling in the plane and reflecting off the boundary of the billiard table. The program records the boundary points the trajectory hits, stopping after a specified number of collisions. The program then produces a visual of the billiard trajectory.

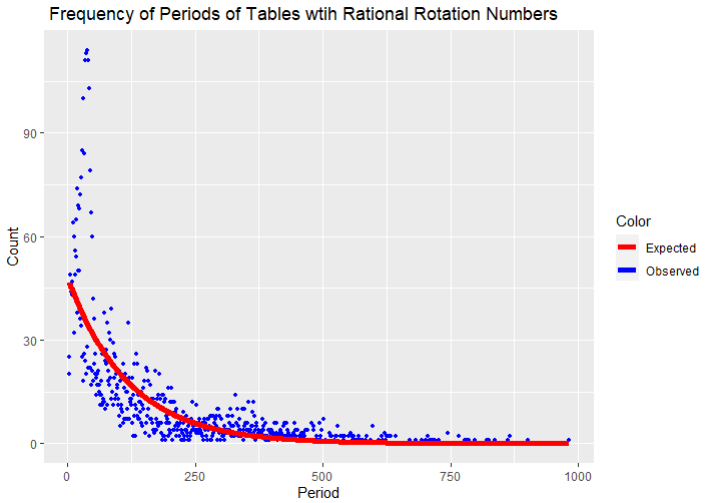
For the algebraic program, we applied the general billiard rule $g$ for unusual billiards described in section 2.3.2. The goal of this program was to find the periods of periodic orbits for different choices of $\rho_1$ and $\rho_2$. A choice of rotation numbers is made, along with an initial point on the caustic $a_0 \in C_0$. The program then produces a sequence of points on the caustic using the rule $a_{n+1} = g(a_n)$ stopping when the sequence returns to the initial point (note that internally the program using the $p, q$ coordinates described in section 2.3.2). The program then returns the observed period of the trajectory. That is, it returns the number of times $g$ was applied to the initial point to complete one orbit. A function in the program runs this process for many choices of initial points given two fixed rotation numbers. This function records the observed periods for each trajectory and writes them to a set. This allows us to create a dataset with the independent variable being the choice of rotation numbers and the dependent variable being the observed periods.

## 3.1 Statistical Analysis

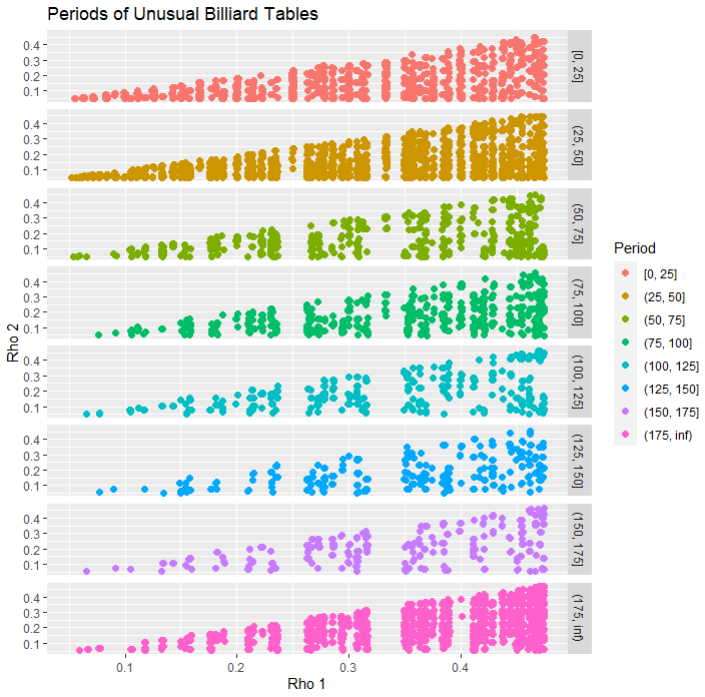
This analysis was conducted in R. The initial dataset comprised of the periods from tables with each semi-circle having rational rotation numbers with denominators ranging from 1 to 10. When the period is treated as a random variable then the frequency a certain period appears seems to follow a Geometric distribution.

Operating with $\hat{\theta} = \frac{1}{\bar{x}}, \alpha = 0.05$, we failed to reject the null hypothesis that "the period follows a Geometric distribution".

When the dataset is expanded to rational rotation numbers with denominators ranging from 1 to 20, and operating under the same hypothesis and assumptions as the first test, we reach the same conclusion and fail to reject the null hypothesis.




Next, the rotation numbers were plotted along the $x$ and $y$ axes and the periods categorized into into bins with widths of 25. There are a few details to note. First, the bin width was chosen arbitrarily at 25 and further research is needed to determine a more optimal choice. Second, each of the graphs only has points populating the bottom right triangle because of the restriction that $\rho_1 < \rho_2$.




The main item of interest and aspect for further research is the clear defined sections that appear in the graph. The two most apparent sections are when $\rho_1 = 0.25, \rho_1 \approx 0.32$, regardless of $\rho_2$, the periods have an upper limit of 100 when $\rho_1 \approx 0.32$ and 50 when $\rho_1 = 0.25$. In addition, there is a lack of observations around those two points, however this could simply be due to having a limited data set and range of values. Furthermore, as the period increases there are increasingly more regions without any observations.

# References

[DR15] Vladimir Dragović and Milena Radnović. *Periods of Pseudo-Integrable Billiards*. 2015. DOI: 10.1007/s40598-014-0004-0. URL: https://link.springer.com/article/10.1007/s40598-014-0004-0.

[DR12] Vladimir Dragović and Milena Radnović. *Pseudo-integrable billiards and arithmetic dynamics*. 2012. DOI: 10.48550/ARXIV.1206.0163. URL: https://arxiv.org/abs/1206.0163.

[Tab95] Serge Tabachnikov. *Billiards*. 1st ed. American Mathematical Society, 1995, pp. 4–21.

[Wat82] Peter Watlers. *An Introduction to Ergodic Theory*. 1st ed. New York, NY: Springer, 1982.