SEQUENCES OF BILLIARD BALL COLLISIONS

JONATHAN ALLEN, JOHN WANG

Abstract.

1. Introduction

In this paper we will explore sequences of billiard ball collisions. In particular, we look at the sequence of sides that a billiard ball collides with under perfect, frictionless conditions. We will show how a square billiard table can be analyzed by tiling the table in the plane, and prove a number of properties that billiard ball sequences must satisfy.

This introduction will explain the general setup of the problem and will give an example of how the definitions relate to a billiard ball with particular initial conditions.

1.1. **Setup.**

We will imagine an infinitesimally small billiard ball on a square table. For simplicity, we will assume that the square table is defined on the unit square $[0,1]^2$. The ball will start at some initial position $\mathbf{x}_0$ inside of the table and with some velocity $\mathbf{v}_0$. We will assume that the ball is massless and frictionless, and that there is no gravity.

We will assume ideal, elastic collisions. To be more precise, when the ball collides with an edge of the table, the ball's velocity will be reflected across the line perpendicular to the edge of the table at the point of collision. In other words, the angle of incidence is equal to the angle of reflection on all billiard ball collisions. Figure 1 shows the general mechanics of a collision.

Now, we will label the horizontal sides of the table h and the vertical sides of the table v. Whenever the billiard ball collides with a horizontal side (labelled h), we will call the resulting collision an h-collision. Whenever the ball collides with a vertical side (labelled v), we will call the resulting collision a v-collision.

We will now define what this paper will be primarily interested in, a sequence of collisions:

Date: December 5, 2013.

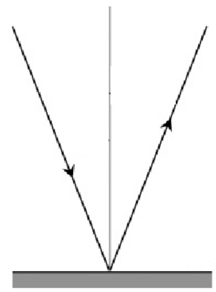

FIGURE 1. Mechanics of a Billiard Ball Colliding with a Table Edge

Definition 1.1. *A* collision sequence *(α) is a sequence of v and h collisions which starts and ends with an h collision for some ball b with initial position $\mathbf{x}_0$ and intial velocity $\mathbf{v}_0$.*

Notice that all non-trivial initial conditions for a billiard ball will result in infinitely many h-collisions. The only initial conditions for which this is not true are when the initial velocity is parallel to the horizontal ($\mathbf{v}_0 = (1, 0)$) so that the ball bounces infinitely between the two vertical sides. The proof of this is trivial and should become clear once the tiling representation is presented, so we will omit it.

Thus, it is perfectly reasonable to constrain a ball's collision sequence to begin and end with an h collision, since one simply needs to extend the number of collisions one watches until the sequence of collisions begins and ends with an h-collision. This constraint will later make it easier to reason about properties of sequences.

1.2. **Example Collision Sequence.** To understand collision sequences better, we will provide an example. Consider a billiard ball with intial position $\mathbf{x}_0 = (0.75, 0.75)$ and initial velocity $\mathbf{v}_0 = (0.23, 0.05)$.

We can see from the figure that the collisions that it makes, denoting a v-collision with a v and an h-collision with an h, are $\{v, h, v, v, v, v, v, h, v, v, v, v, v, h, v, v, v\}$. A valid collision sequence given these initial conditions would start and end with an h collision. A possible example would be:

$$\alpha = hvvvvvhvvvvvh \tag{1}$$

Notice that the collisions will continue infinitely, so one could imagine extending this collision sequence to include more v and h collisions. Although not shown in the picture, the following collision sequence

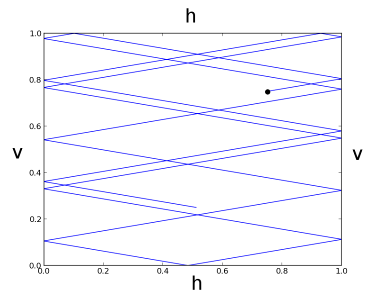

FIGURE 2. Example Billiard Ball Trajectory, $x_0 = (0.75, 0.75)$, $v_0 = (0.23, 0.05)$

would also be valid if one continued showing the trajectory of the ball in future collisions:

$$\alpha = hvvvvvhvvvvvhvvvvvh \tag{2}$$

1.3. **Simplifications.** There are also a couple of simplifications that we can make without loss of generality that will make talking about billiard balls, their initial conditions, and their collision sequences easier.

- We can change the magnitude of the initial velocity $\mathbf{v}_0$ without changing any collision sequences. This is because only the direction of $\mathbf{v}_0$ affects the points of collision of the billiard ball with the table. Thus, the only part of the initial velocity that we care about is the velocity's direction. We could alternatively talk about the angle γ that $\mathbf{v}_0$ makes with the positive horizontal line (just as in polar coordinates).
- We can constrain the initial velocity's angle to the range $\gamma \in [0, \pi/2]$. This is because any intial velocity in the range $\gamma \in [-\pi/2, \pi/2]$ will create equivalent collision sequences as $\pi + \gamma$ (this will be clear once we present the tiling representation later in the paper). Moreover, any angle in the range $\gamma \in [0, \pi/2]$ will create equivalent collision sequences as those in $[-\pi/2, 0]$ by simply reflecting the cube about one of its horizontal sides.

Using these simplifications, we will constrain the initial conditions of all billiard balls throughout the rest of the paper so that the initial velocity $\mathbf{v}_0$ has an angle $\gamma \in [0, \pi/2]$.

2. Tiling Representation

We will now present a representation of the problem which will greatly simplify the analysis of v and h collisions for some ball B, called the tiling representation. First we will define some basic definitions we need to explain the tiling representation.

Definition 2.1. *A billiard ball* trajectory $\tau(t_0, t_1)$ *is the curved traced by a billiard ball B between times t_0 and t_1.*

Definition 2.2. *A collision time κ_i for a billiard ball B is the time at which the ith collision occurs.*

To understand the basics of how the tiling representation works, imagine placing a square billiard table on the xy plane. The billiard table is the unit square$[0,1]^2$. The table's edges will be the four line segments bordering the unit square. A ball will start with some initial position $\mathbf{x}_0 \in [0,1]^2$ and velocity $\mathbf{v}_0$. After some time, the ball will collide with an edge e_0 of the table at time κ_1. However, instead of thinking of the trajectory of the ball as being reflected across the line perpendicular to e_0 at the point of collision, we will instead reflect the original unit square s_0 across the edge e_0 to create a new square s_1. Now, the trajectory $\tau(\kappa_1, \kappa_2)$ of the ball after the first collision will be traced in the new square s_1.

In other words, the trajectory $\tau(0, \kappa_1)$ before the first collision will be confined to the original square s_0, and the trajectory$\tau(\kappa_1, \kappa_2)$ after the first collision will be confined to the new reflected square s_1. We can continue the process for each new collision. Suppose the ball collides with edge e_1 in square s_1. Then, we will create a new square s_2 which is a reflection of square s_1 across the edge e_1. The trajectory of the ball $\tau(\kappa_2, \kappa_3)$ after the second collision will be confined to the newest reflected square s_2. This process will continue on indefinitely so that the trajectory $\tau(\kappa_j, \kappa_{j+1})$ will be confined to the square s_j, where square s_{j+1} is generated by reflecting square s_j across the edge e_j which is collided with at time κ_j.

Figure 3 shows an example trajectory which is created using this tiling process. In essence, the tiling representation reflects a table about each of its four sides. These reflections will perform the same process, eventually tiling and completely filling the xy plane. The trajectory of particular billiard ball can then be traced through the tiling in the xy-plane, as seen in figure 3.

A couple of interesting observations can be made:

- The combined trajectory $T = \{\tau(0, \kappa_1), \tau(\kappa_1, \kappa_2), \tau(\kappa_2, \kappa_3), \ldots\}$ of a ball creates a ray in the xy plane.

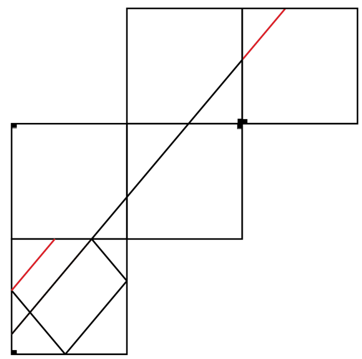

FIGURE 3. Example of Tiling Billiard Tables

- All v-collisions happen exactly when the combined trajectory T intersects with the integer vertical lines $x = k$ where $k \in \mathbb{Z}$. The same can be said of h-collisions and integer horizontal lines $y = k$ for $k \in \mathbb{Z}$.

We can formalize and prove each of these observations in turn:

Theorem 2.3. *The combined trajectory* $T = \{\tau(0, \kappa_1), \tau(\kappa_1, \kappa_2), \ldots\}$ *of a billiard ball is a ray in the xy plane under the tiling representation.*

Proof. We need to show that all trajectories $\tau(0, \kappa_1), \tau(\kappa_1, \kappa_2), \ldots$ which constitute the combined trajectory lie on a single line. We know that trajectories $\tau(\kappa_i, \kappa_{i+1})$ are line segments because the velocity of the billiard ball only changes during a collision. Thus, we just need to show that each trajectory $\tau(\kappa_i, \kappa_{i+1})$ lies on the same line, i.e. that $\tau(\kappa_i, \kappa_{i+2})$ is a line segment for all $i \in \mathbb{Z}$.

We will show this by analyzing the change in trajectory at time κ_i. We know that at any time $t \in (\kappa_{i-1}, \kappa_i)$, the billiard ball will be on the line segment defined by the trajectory $\tau(\kappa_{i-1}, \kappa_i)$. This trajectory makes an angle θ with the edge e_i which the billiard ball collides with at time κ_i. Moreover, we know that after colliding with e_i, the outgoing angle of the trajectory is equal to the incoming angle θ by our definition of collision.

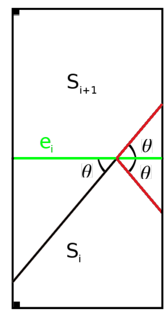

FIGURE 4. Trajectories $\tau(\kappa_{i-1}, \kappa_i)$ in s_i and $\tau(\kappa_i, \kappa_{i+1}$ s_{i+1} forming a line segment

The velocity is reflected about the line perpendicular to e_i at the point of collision, but the angle of the trajectory $\tau(\kappa_i, \kappa_{i+1})$ makes with e_i remains the same. Thus, when square s_i is reflected about e_i, the resulting trajectory in s_{i+1} makes an angle of θ with e_i. Figure 4 shows this process graphically.

It is clear that the angles that $\tau(\kappa_{i-1}, \kappa_i)$ and $\tau(\kappa_i, \kappa_{i+1}$ make with e_i are both θ. Since both of these trajectories are already line segments, we see that the union of the two trajectories is also a line segment because both trajectories lie on the same line. Thus, we see that all trajectories $\tau(\kappa_j, \kappa_{j+1})$ for all $j \in \mathrm{Z}$ lie on the same line, which completes the proof. $\square$

Theorem 2.4. *Let $t_{v,k}$ be a time when the combined trajectory T intersects with some integer vertical line $x = k$ where $k \in \mathrm{Z}$. We must have $\kappa_k^v = t_{v,k}$, where κ_k^v is the time at which the kth v-collision occurs.*

Similarly, let $t_{h,k}$ be a time when the combined trajectory T intersects with some horizontal vertical line $y = k$ for $k \in \mathrm{Z}$. Then, $\kappa_k^h = t_{h,k}$.

Proof. The proof of the first statement is exactly analogous to the proof of the second statement, so we will only prove the theorem for $\kappa_k^v = t_v$.

Recall that in the tiling representation, whenever the billiard ball collides with an edge e_j, a new square s_{j+1} gets reflected across e_j. It is clear that e_j, if it is a horizontal edge, lies on some integer horizontal line $y = c$ where $c \in \mathrm{Z}$. Alternatively, if e_j is a vertical edge then it lies on some integer vertical line $x = c$ where $c \in \mathrm{Z}$. Thus each κ_j corresponds to when the combined trajectory T intersects with some integer vertical or integer horizontal line.

It is also clear that the billiard ball cannot make a v-collision at time t unless the combined trajectory T intersects with an integer vertical line at time t as well. Therefore, we see that v-collisions occur exactly when T intersects with an integer vertical line.

We know that T traces a ray in the xy plane by theorem 2.3. By assumption, this ray starts in the unit square $[0,1]^2$ and has an angle between 0 and $\pi/2$ with the horizontal. Thus, the first v-collision occurs when T intersects $x = 1$, the second v-collision occurs when T intersects $x = 2$, and the kth v-collision occurs when T intersects $x = k$.

However, we know that the kth v-collision occurs at time κ_k^v by definition, and that the $t_{v,k}$ is the time when T intersects with $x = k$. Therefore, we see that $\kappa_k^v = t_{v,k}$. □

We now see that we can represent the combined trajectory T of a billiard ball as a ray in the plane. We can define the line which the ray lies on as $y = mx + y_0$, where m is the slope of the line and is given by $m = \frac{\mathbf{v}_y}{\mathbf{v}_x}$ and y_0 is the y-intercept which can be determined by $\mathbf{x}_0$ by solving for $\mathbf{x}_y = m\mathbf{x}_x + y_0$. This line represents the entire combined trajectory, and also determines all possible collision sequences for a particular billiard ball, since v-collisions happen when x is an integer and h-collisions happen when y is an integer.

3. Simple Properties

We will now use the tiling representation that we have developed to discover properties of collision sequences. The first simple property is that collision sequences are periodic when the initial conditions are rational numbers. Now that we can define a billiard ball as a line, instead of giving initial conditions $\mathbf{x}_0$ and $\mathbf{v}_0$, we can give the slope m and the y-intercept y_0 of the complete trajectory's line. The formalized theorem is then:

Theorem 3.1. *There exists a $k \in \mathrm{N}$ such that $y_0 \equiv mk + y_0 \pmod 2$ if and only if $m \in \mathrm{Q}$.*

Proof. Let us first show that if $m \in \mathrm{Q}$, then there exists a $k \in \mathrm{N}$ such that $y_0 \equiv mk + y_0 \pmod 2$. We simply need to show that $0 \equiv mk \pmod 2$ if $m \in \mathrm{Q}$. However, since we know $m \in \mathrm{Q}$, we can decompose it as follows $m = p/q$ where $p, q \in \mathrm{Z}$. Thus, we have $mk \pmod 2 \equiv \frac{pk}{q} \pmod 2$. Now we can choose:

$$k = \begin{cases} q & \text{if } p \pmod 2 \equiv 0 \\ 2q & \text{if } p \pmod 2 \equiv 1 \end{cases} \tag{3}$$

In this way, we see that $0 \equiv \frac{pk}{q} \pmod 2$, which proves the first half of the theorem.

Now let us show that if there exists a in $k \in \mathrm{N}$ such that $y_0 \equiv mk+y_0 \pmod 2$, then $m \in \mathrm{Q}$. If such a k exists, then we must have $0 \equiv mk \pmod 2$, which means that $mk = 2q$ for some $q \in \mathrm{Z}$. This means $m = \frac{2q}{k}$. Now it is clear that $m \in \mathrm{Q}$ because both its numerator and denominator are integers. □

Theorem 3.1 shows that if $m \in \mathrm{Q}$, then the billiard ball will eventually return to it's original position $\mathbf{x}_0$ with its original velocity $\mathbf{v}_0$. Seeing why this is true is just a matter of using the tiling representation. We note that every second square in either the x or y direction is the same (because of the transitivity of reflection). Therefore, every second square will have a trajectory that exactly corresponds to the trajectory in the original square.

Thus, if $y_0 \equiv mk + y_0 \pmod 2$, then the y-intercept from one of the secondary squares is the same as the y-intercept from the original square. Since we know that the secondary squares have the same trajectories as in the original square, we see that the trajectory will have returned to its original position (since the velocity is the same). Thus, if $y_0 \equiv mk + y_0 \pmod 2$, then we know that the billiard ball will return to its original position with its original velocity. Theorem 3.1 also shows that if m is irrational, then the billiard ball will never return to its original position and velocity.

We can also examine consecutive occurences of v and h. For example, can we have consecutive occurences of both v and h like in the sequence $hhvvhh$? In fact, we cannot as theorem 3.2 shows.

Theorem 3.2. *A valid collision sequence cannot have consecutive occurrences of v and consecutive occurences of h.*

Proof. We have already shown in theorem 2.3 that the combined trajectory of a billiard ball must be a ray. This ray must lie on some line with some slope $m \in \mathrm{R}$ or with undefined slope (when the line is vertical). When the line is vertical, it is clear that the theorem holds, because only v collisions occur.

There are two cases left: $|m| < 1$ or $|m| \geq 1$. If $|m| \geq 1$, then the number of v-collisions between the xth and the $x+1$st h-collision will be $\lfloor y(x+1) - y(x) \rfloor = \lfloor m(x+1) + y_0 - (mx + y_0) \rfloor = \lfloor m \rfloor \geq 1$. Thus, we see that for any $x \in \mathrm{N}$, we must have at least 1 v-collision between the xth and $x+1$st h-collisions, which proves the theorem for $|m| \geq 1$.

If $|m| < 1$, then the number of h-collisions between the yth and $y+1$st v-collisions will be $\lfloor x(y+1) - x(y) \rfloor = \lfloor (y+1-y_0)/m - (y-y_0)/m \rfloor = \lfloor 1/m \rfloor > 1$. This shows that there will be at least 1 h-collision between the yth and $y+1$st v-collisions for all $y \in \mathrm{N}$. This completes the theorem. □

We therefore see that if v occurs consecutively in a collisions sequence, then h cannot occur consecutively and vice versa. For example, the sequence *hvvhh* has two consecutive occurrences of v and two consecutive occurrences of h, so it cannot be a valid collision sequence. However, the sequence *hvvhvvvh* does not have consecutive occurrences of h, so theorem 3.2 does not rule it out as a valid collision sequence.

Theorem 3.2 allows us to make a simplification for our collision sequences. Since one of either v or h must occur non-consecutively, we can abitrarily assign h to be the side that occurs non-consecutively by rotating the billiard table and creating the opposite collision sequence. For example, the sequence *vhhhvhhhv* is the same as the sequence *hvvvhvvvh* when one rotates the billiard table by $\pi/2$. Therefore, from now on, we can confine all our sequences to have only non-consecutive occurrences of h (i.e. sequences like *vhhhvhhhv* become *hvvvhvvvh*) without loss of generality.

4. 1-Dimensional Representation

Note: There are more h's than v's: $m \geq 1$

Rather than looking at an explicit representation of lines in the plane, we can gain much more insight from looking at a parametric representation. To simplify our analysis, we will choose our time parameter such that v collisions occur every $\Delta t = 1$ and h collisions occur every $\Delta t = m$. The equation for a line $y(x) = m\,x + b$ is equivalent to the following parametric system

$$x(t) = \frac{1}{m}\,t + x_0 \tag{4}$$

$$y(t) = t \tag{5}$$

Now our v and h collisions in the 2-dimensional plane can be projected onto the 1-dimensional t axis.

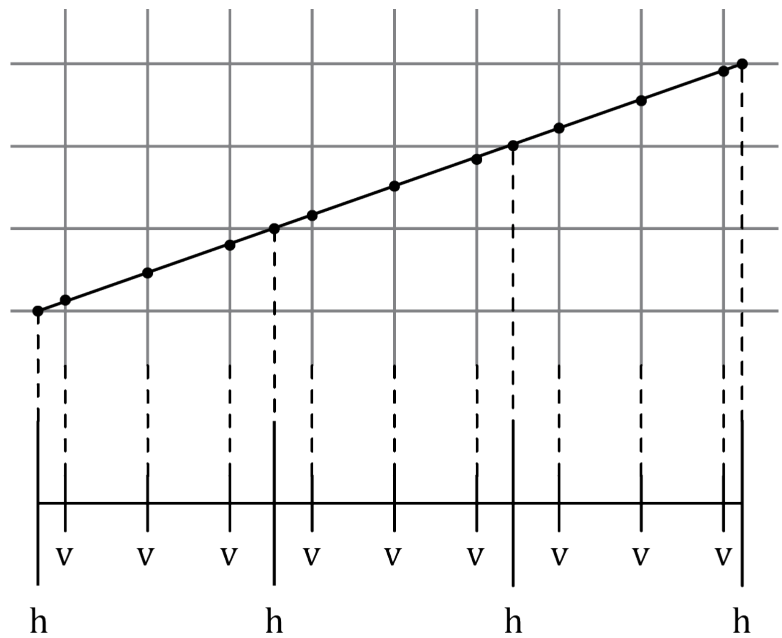

FIGURE 5. Projecting onto the parametric representation.

Lemma 4.1. *A sequence α is a valid collision sequence iff there exists at least one valid collision sequence containing α that starts and ends with an h.*

Because of Lemma 4.1, we will confine ourselves to only looking at collision sequences that start and end with an h.

Definition 4.2. *Given a collision sequence α, define a sequence β where each element β_i is the number of v collisions between the i^{th} and $(i+1)^{th}$ h in α.*

The β sequence is much simpler to think of geometrically in terms of our parametric representation shown in Figure 5. β_i represents the number of v collision tick marks in between each h collision tick mark.

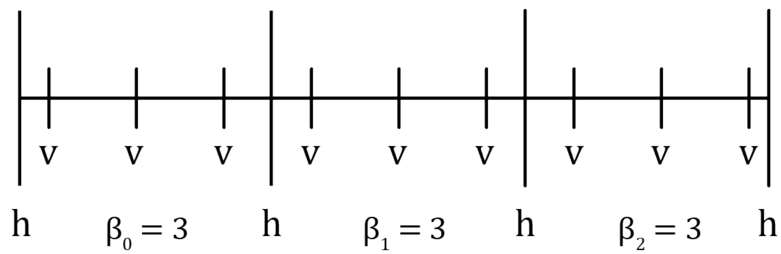

FIGURE 6. The β sequence.

Lemma 4.3. *For every valid collision sequence, the following must be true*

$$\beta_{min} > 0 \tag{6}$$

Proof. TODO □

Theorem 4.4. *For every valid collision sequence, the following must be true*

$$\beta_{max} - \beta_{min} \leq 1 \tag{7}$$

Proof. From Equation 4, v collisions occur every $\Delta t = 1$ and h collisions occur every $\Delta t = m$. Thus, the following must be true

$$\beta_i \in (\lfloor m \rfloor, \lceil m \rceil) \tag{8}$$

For an m to exist that satisfies the above constraints, all numbers in the β sequence can only differ by 1.

□

From now on we will only consider collision sequences, where β starts and ends with $\beta_m in$.

Theorem 4.5. *Define*

$$\delta_i^{(\beta)} := \begin{cases} x_0 & \textit{if} \quad i = 0 \\ i(\beta_{max} - m) & \textit{otherwise} \end{cases} \tag{9}$$

Then the following is true for all valid collision sequences

$$\beta_i = \left\lfloor \delta_i^{(\beta)} \right\rfloor + \beta_{max} - \left\lfloor \delta_{i+1}^{(\beta)} \right\rfloor \tag{10}$$

Proof. TODO... □

We can immediately notice that the $\delta^{(\beta)}$ sequence has the following features:

(1) The $\delta^{(\beta)}$ sequence is increasing, because $\beta_{max} \geq m$
(2) Combining Theorem 4.4 and Equation 10, we get the following:

$$\left\lfloor \delta_{i+1}^{(\beta)} \right\rfloor - \left\lfloor \delta_{i+1}^{(\beta)} \right\rfloor = \beta_{max} - \beta_{min} \tag{11}$$

$$\leq 1 \tag{12}$$

Thus, if we plot the values of the $\delta^{(\beta)}$ sequence on a line, we notice something interesting: the plot looks very similar to our original plot of the collision sequence parameterized by t.

Definition 4.6. *Define a the sequence $C^{(0)}$ where each element $C_i^{(0)}$ is 1 more than the number of occurrences of β_{max} between the i^{th} and (i+1)th occurrence of β_{min} in the β sequence.*

Theorem 4.7. *Define*

$$\delta_i^{(C^{(j)})} := \begin{cases} x_0 & \text{if} \quad i = 0 \\ i(C_{max}^{(j)} - m) & \text{otherwise} \end{cases} \tag{13}$$

Then the following is true for all valid collision sequences

$$C_i^{(j)} = \left\lfloor \delta_i^{(C^{(j)})} \right\rfloor + C_{max}^{(j)} - \left\lfloor \delta_{i+1}^{(C^{(j)})} \right\rfloor \tag{14}$$

Theorem 4.8. *Define the sequence a as*

$$a_0 := 1 \tag{15}$$

$$a_1 := \beta_{max} - m \tag{16}$$

$$a_i := C^{(i-2)} a_{i-1} - a_{i-2} \qquad \text{for} \quad i \geq 2 \tag{17}$$

For every valid collision sequence, $a \to 0$

Proof. □

5. Conclusion

TODO