SEQUENCES OF BILLIARD BALL COLLISIONS

JONATHAN ALLEN, JOHN WANG

Abstract. In this paper we look at sequences of collisions that a billiard ball undergoes when bouncing around inside a billiard table. We present a tiling representation to help elucidate properties of these sequences. We are then able to relate this tiling representation to a simpler one-dimensional problem, that allows us to completely characterize the sequences through a group of inductively defined subproblems. Finally, we relate these subproblems to continued fractions.

1. Introduction

The problem we will analyze in this paper is that of a billiard ball bouncing around inside a square billiard table in the absence of any external forces (e.g. gravity) and/or any dissipative forces (e.g. friction). Before delving into the analysis, we must first clarify the problem statement.

1.1. Setup.

The ball and table will be idealized in the following manner: we will represent the ball as a point moving around in $\mathbb{R}^2$ and the board as the unit square $[0,1]^2$. The ball will start with an initial position $\mathbf{x}_0 \in [0,1]^2$ and initial velocity $\mathbf{u}_0$. Because there are no external forces and/or dissipative forces, the speed of the ball is constant and irrelevant to the problem.

Any time the position of the ball (point) coincides with the edge of the table (unit square), we will say that the ball collides with that edge of the table. During this collision, the ball is reflected off the edge of the table in such a manner that the outgoing direction vector is a reflection of the incoming direction vector across the line perpendicular to the edge of the table at the point of collision. In more technical terms, the angle of incidence is equal to the angle of reflection in all ball-table collisions. Collisions at corners of the table are undefined and so we will ignore any such trajectories that intersect corners of the table. Figure 1 shows the general mechanics of a collision.

Date: December 12, 2013.

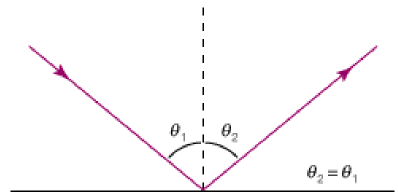

FIGURE 1. Mechanics of the ball colliding with a table edge.

Introducing some notation to the problem, we will label the horizontal edges of the table h and the vertical edges of the table v. Whenever the ball collides with a horizontal edge, we will call the resulting collision an h collision. Likewise, collisions with vertical edges will be denoted as v collisions.

Definition 1.1. *A* collision sequence *(α) for a ball is the sequence of sides that the ball collides with ($\alpha_i \in \{v, h\}$). This sequence is ordered by increasing collision time. In this paper, ball trajectories are idealized and infinite, but we will only look at finite subsequences of the infinite collision sequences formed by these trajectories. Furthermore, we will only look at finite collision sequences that start and end with h.*

1.2. **An Example Collision Sequence.** For the reader to better understand the basic properties of collision sequences, we will consider an example trajectory and form a collision sequence from a short segment of the trajectory. Consider a ball with initial position $\mathbf{x}_0 = (0.75, 0.75)$ and initial velocity $\mathbf{u}_0 = (4.6, 1)$.

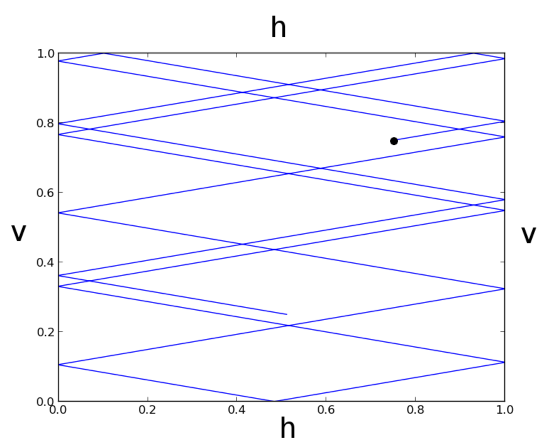

FIGURE 2. Example trajectory, $x_0 = (0.75, 0.75)$, $u_0 = (4.6, 1)$.

The collision sequence corresponding to this trajectory segment is the following:

$$\alpha = (v, h, v, v, v, v, v, h, v, v, v, v, v, h, v, v, v) \tag{1}$$

2. Tiling Representation

We can now dive into analyzing the problem. We will start off by introducing a tiling which will greatly simplify the analysis of trajectories and collision sequences. Before doing so, we need to introduce the following terminology:

Definition 2.1. *A ball's* trajectory $\tau(t_0, t_1)$ *is the piecewise linear path traced by the ball between times* t_0 *and* t_1*.*

Definition 2.2. *A collision time* κ_i *for a ball is the time at which the* i^{th} *collision occurs.*

To understand the basics of how the tiling representation works, imagine placing a unit square table on the xy plane. 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 $\mathbf{u}_0$. 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, as seen in figure 4.

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 combined trajectory of a particular ball can then be traced through the tiling in the xy-plane, as seen in figure 3.

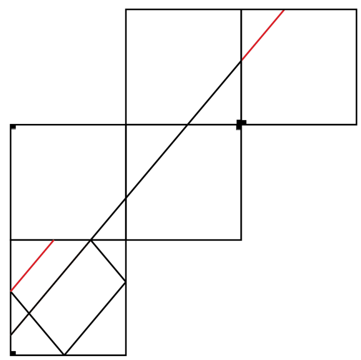

Figure 3. Example of Tiling Billiard Tables.

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.
- 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 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 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 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 ball collides with at time κ_i. Moreover, we know that after colliding with e_i, the angle of reflection of the trajectory is equal to the angle of incidence θ by our definition of collision.

The velocity is reflected about the line perpendicular to e_i at the point of collision, but the angle that 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.

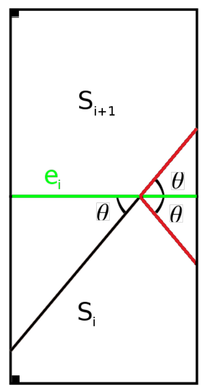

FIGURE 4. Trajectories $\tau(\kappa_{i-1}, \kappa_i)$ in s_i and $\tau(\kappa_i, \kappa_{i+1}$ s_{i+1} forming a line segment.

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. 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. □

We will now show that all v-collisions happen exactly when the combined trajectory T intersects with the integer vertical lines $x = k$ where $k \in \mathrm{Z}$, and all h-collisions happen when the combined trajectory T intersects with the integer horizontal lines $y = k$ for $k \in \mathrm{Z}$. To make this notion precise, we will define times at which these intersections occur in theorem 2.4. We will show that the time of intersection between the combined trajectory and a vertical line must equal the time of collision between the ball and a v side, and an analogous result for horizontal lines.

Theorem 2.4. *Let $t_{v,k}$ be a time when the combined trajectory T for some ball B 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* $k \in \mathbb{Z}$. *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 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 $c \in \mathbb{Z}$. Alternatively, if e_j is a vertical edge then it lies on some integer vertical line $x = c$ where $c \in \mathbb{Z}$. 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 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 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 $m = \frac{\mathbf{u}_y}{\mathbf{u}_x}$, where $\mathbf{u}_y$ and $\mathbf{u}_x$ are the y and x components of $\mathbf{u}_0$ respectively. Also, y_0 is the y-intercept which can be determined using $\mathbf{x}_0$ by solving for $\mathbf{x}_y = m\mathbf{x}_x + y_0$, where $\mathbf{x}_y$ and $\mathbf{x}_x$ are the y and x components of $\mathbf{x}_0$ respectively. The resulting line $y = mx + y_0$ represents the entire combined trajectory, and also determines all possible collision sequences for a particular ball, since v-collisions happen when x is an integer and h-collisions happen when y is an integer.

Thus, we see that the ith v-collision occurs when $y = mi + y_0$ and the ith h-collision occurs when $x = (i - y_0)/m$.

3. Simple Properties

We will now use the tiling representation that we have developed to derive some simple properties of all valid collision sequences. The first property is that collision sequences are periodic when the components of the initial direction of motion vector are rational numbers. Now that we can equate a ball's trajectory with a line, instead of defining the trajectory by its initial conditions $\mathbf{x}_0$ and $\mathbf{u}_0$, we can define it by the slope m and the initial y-intercept, $y_0 := y(0)$, of the complete trajectory's line.

Theorem 3.1 will show that if $m \in \mathrm{Q}$, then the ball will eventually return to it's original position $\mathbf{x}_0$ with its original direction of motion $\mathbf{u}_0$. 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, when analyzing trajectories that are periodic, we must work modulo 2, since it takes two reflections in order to "straighten out" the table's reflection and get the original table's orientation back.

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 ball will return to its original position with its original velocity. Theorem 3.1 also shows that if m is irrational, then the ball will never return to its original position and velocity.

The formalized theorem is:

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 into $m = p/q$ where $p, q \in \mathrm{Z}$. Thus, we have $mk \pmod 2 \equiv \frac{pk}{q} \pmod 2$. Now we can choose:

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

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 $0 \equiv mk$

(mod 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. $\square$

We can also examine consecutive occurrences of v and h. For example, can we have consecutive occurrences of both v and h like in the sequence $\alpha = (h, h, v, v, h, h)$? In fact, we cannot as theorem 3.2 shows.

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

Proof. We have already shown in theorem 2.3 that the combined trajectory of a 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 $\lfloor (m(x+1)+y_0) - (mx+y_0) \rfloor = \lfloor m \rfloor \geq 1$. 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 $\lfloor ((y+1)-y_0)/m - (y-y_0)/m \rfloor = \lfloor 1/m \rfloor > 1$. 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. $\square$

We therefore see that if v occurs consecutively in a collision sequence, then h cannot occur consecutively and vice versa. For example, the sequence $\alpha = (h, v, v, h, h)$ has two consecutive occurrences of v and two consecutive occurrences of h, so it cannot be a valid collision sequence. However, the sequence (h, v, v, h, v, v, v, h) 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 arbitrarily assign v to be the side that occurs non-consecutively by rotating the table and creating the opposite collision sequence. For example, the sequence $(v, h, h, h, v, h, h, h, v)$ is the same as the sequence $(h, v, v, v, h, v, v, v, h)$ when one rotates the table by $\pi/2$. Therefore, from now on, we can confine all our sequences to have only non-consecutive occurrences of v (i.e. sequences like $(h, v, v, v, h, v, v, v, h)$ become $(v, h, h, h, v, h, h, h, v)$) without loss of generality.

4. 1-Dimensional Representation

Rather than looking at the explicit equation for a line in the plane, we can gain much more insight from looking at the parametric equation parameterized by t. To simplify our analysis, we will choose our parameter t such that v collisions occur every $\Delta t = 1$ and h collisions occur every $\Delta t = m$. The equation for a line $y(x) = m\,x + y_0$ is equivalent to the following parametric system

$$x(t) = \frac{1}{m}\,t \tag{3}$$

$$y(t) = t + y_0 \tag{4}$$

Now v and h collisions in the 2-dimensional plane can be projected onto the 1-dimensional t axis as is shown in Figure 5.

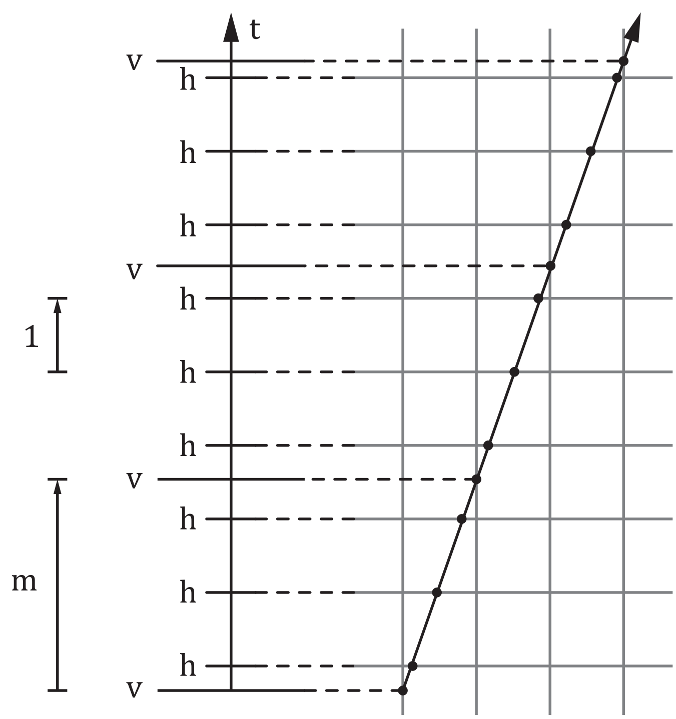

Figure 5. Projecting onto the t axis.

For the sake of space, we will rotate the t axis so that it is horizontal, as is shown in Figure 6. Thus, the problem of mapping collision sequences to lines in the plane becomes a problem of fitting regularly spaced tick marks into intervals on the t axis.

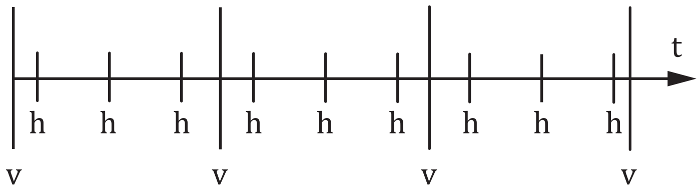

FIGURE 6. The 1-dimensional problem.

To start off analyzing this 1-dimensional problem, we need a basic lemma about counting ticks marks inside an interval.

Lemma 4.1. *Let $l_2 < l_1$. The number of real numbers at regular spacing l_2 inside an open interval of length l_1 is in $\left\{\left\lfloor \frac{l_1}{l_2} \right\rfloor, \left\lceil \frac{l_1}{l_2} \right\rceil\right\}$.*

Proof. Let $A \subset \mathbb{Z}$ represent the set of all possible numbers of real numbers at regular spacing l_2 inside an open interval of length l_1. Given some $a \in A$, let $b \in \mathbb{R}$ be defined such that the following holds

$$l_1 = (a-1)l_2 + b \tag{5}$$

We know that $(a-1)l_2 < l_1$, so $b \geq 0$. Also, $b < 2\,l_2$, because otherwise, there must be more than a real numbers inside the interval, contradicting our original assumption.

Rearranging Equation 5, we get

$$a = \frac{l_1}{l_2} + 1 - \frac{b}{l_2} \tag{6}$$

$$\frac{l_1}{l_2} - 1 < a < \frac{l_1}{l_2} + 1 \tag{7}$$

$$a \in \left\{\left\lfloor \frac{l_1}{l_2} \right\rfloor, \left\lceil \frac{l_1}{l_2} \right\rceil\right\} \tag{8}$$

□

5. Some Useful Sequences

To analyze the 1-dimensional problem, we will derive two groups of sequences from the original collision sequence, which we will call the β and δ groups. These groups will allow us to validate any collision sequence. The first sequence in the β group is defined below.

Definition 5.1. *Given a collision sequence α, define a sequence $\beta^{(0)}$ such that each element $\beta_i^{(0)}$ is the number of h's between the i^{th} and $(i+1)^{th}$ v in α. From Lemma 4.1, each element in $\beta^{(0)}$ can be one of two different values, which we will refer to as $\beta_{min}^{(0)}, \beta_{max}^{(0)}$.*

Graphically, $\beta_i^{(0)}$ represents the number of tick marks in each interval. An example of this is shown in Figure 7.

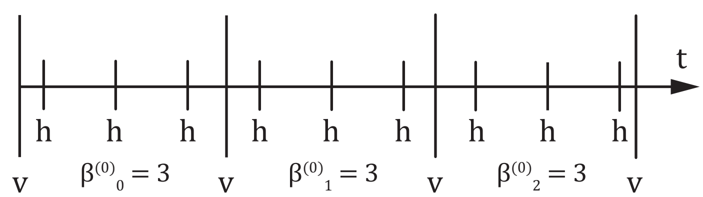

FIGURE 7. An example $\beta^{(0)}$ sequence.

The first sequence in the δ group, $\delta^{(0)}$, is defined such that each element is spaced $\left\{\frac{m}{1}\right\}$ apart. $\delta^{(0)}$ will also include an offset of $-(1-y_0)$ for reasons that will become clear shortly. In terms of our β group, $\left\{\frac{m}{1}\right\}$ can be written as

$$\left\{\frac{m}{1}\right\} = \frac{m}{1} - \beta_{min}^{(0)} \tag{9}$$

Definition 5.2. *Given a $\beta^{(0)}$ sequence, $\delta^{(0)}$ is defined as follows*

$$\delta_i^{(0)} := \begin{cases} -(1-y_0) & \textit{for} \quad i = 0 \\ i(m - \beta_{min}^{(0)}) - (1-y_0) & \textit{for} \quad i \geq 1 \end{cases} \tag{10}$$

Visually, each $\delta_i^{(0)}$ is the distance between the beginning of the i^{th} interval and the $(i\,\beta_{min}^{(0)})^{th}$ tick mark (with positive distance measured right to left). Figure 8 shows the $\delta_i^{(0)}$ sequence on top of the original parametric representation (top of the figure) as well as by itself (bottom of the figure).

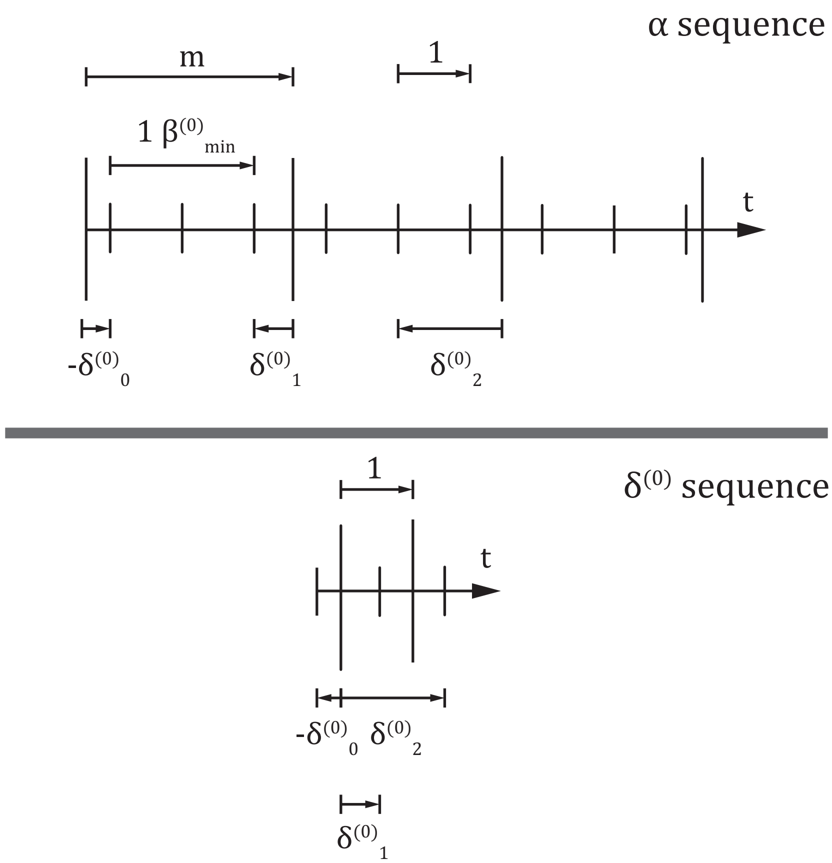

FIGURE 8. Generating the $\delta^{(0)}$ sequence.

Our next step is to define $\beta^{(j)}, \delta^{(j)}$ for $j \geq 1$, which will be done in an inductive manner.

Definition 5.3. *Given a collision sequence α, for some $j > 0$ assume that $\beta^{(j-1)}$ is defined and each element in the sequence is either $\beta^{(j-1)}_{min}$ or $\beta^{(j-1)}_{max}$. The sequence $\beta^{(j)}$ is defined such that each element $\beta^{(j)}_i$ is 1 more than the number of occurrences of $\beta^{(j-1)}_{min}$ between the i^{th} and $(i+1)^{th}$ occurrence of $\beta^{(j-1)}_{max}$ in the $\beta^{(j-1)}$ sequence. From Lemma 4.1, each element in $\beta^{(j)}$ can be one of two different values, which we will refer to as $\beta^{(j)}_{min}, \beta^{(j)}_{max}$.*

If, for some j_f, the length of $\beta^{(j_f-1)}$ is 1, then $\beta^{(j_f-1)}$ is called the terminating β sequence, and all subsequent $\beta^{(j)}$ for $j \geq j_f$ are undefined.

This definition is much easier to explain visually. Figure 12 shows a plot of a $\beta^{(j-1)}$ sequence (the reason for plotting the sequence in this specific manner will be explained later). $\beta^{(j)}$ is formed by counting tick marks in each interval on the t axis.

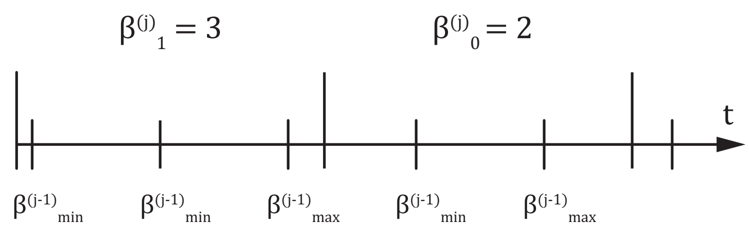

FIGURE 9. A general $\beta^{(j)}$ sequence.

Before continuing on to define $\delta^{(j)}$, let's create a new sequence that will be helpful in the definition.

Definition 5.4. *Given that the β group exists, then a is defined as follows*

$$a_j := \begin{cases} m & \textit{for} \quad i = -2 \\ 1 & \textit{for} \quad i = -1 \\ a_{j-2} - \beta^{(j)}_{min} a_{j-1} & \textit{for} \quad i \geq 0 \end{cases} \tag{11}$$

a_{-2}, a_{-1} are the interval size and tick spacing in our original problem. a_0 is the spacing of elements of $\delta^{(0)}$, which is illustrated in Figure 10.

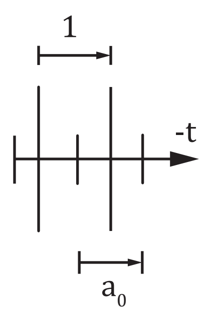

FIGURE 10. a_0.

We can now finish defining the δ group. In words, each $\delta^{(j)}_i$ is the distance between the beginning of the i^{th} interval and the $(i\,\beta^{(j)}_{min})^{th}$ tick mark in the $\delta^{(j-1)}$ plot.

Definition 5.5. $\delta^{(j)}$ *is defined as follows*

$$\delta_i^{(j)} := \begin{cases} -(1-y_0) & \textit{for} \quad i = 0 \\ i(a_{j-2} - \beta_{min}^{(j)} * a_{j-1}) - (1-y_0) & \textit{for} \quad i \geq 1 \end{cases} \tag{12}$$

An example $\delta^{(j-1)}$ sequence is shown in Figure 11.

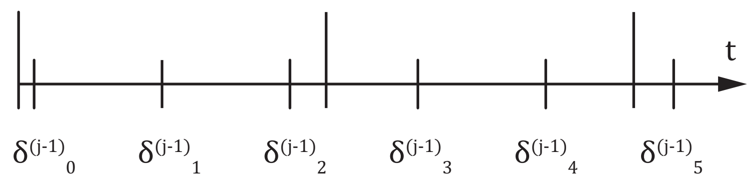

FIGURE 11. An example $\delta^{(j-1)}$ sequence.

To move forward, we need to understand how the $\delta^{(j-1)}$ sequence is related to the $\beta^{(j-1)}$ sequence. Referring back to Figure 8, it can be seen that $\beta_i^{(j-1)}$ is equal to $\beta_{min}^{(j-1)}$ plus the number of tick marks included in the $\delta_{i+1}^{(j-1)}$ interval minus the number of tick marks included in the $\delta_i^{(j-1)}$ interval. More precisely

$$\beta_i^{(j-1)} = \left\lfloor \delta_{i+1}^{(j-1)} \right\rfloor + \beta_{min}^{(j)} - \left\lfloor \delta_i^{(j-1)} \right\rfloor \tag{13}$$

Figure 12 plots the $\beta^{(j-1)}$ sequence in place of the $\delta^{(j)}$ sequence. This plot was shown earlier, but it should be clearer now from Equation 13 why the $\beta_{max}^{(j-1)}$ tick marks appear at the end of each interval.

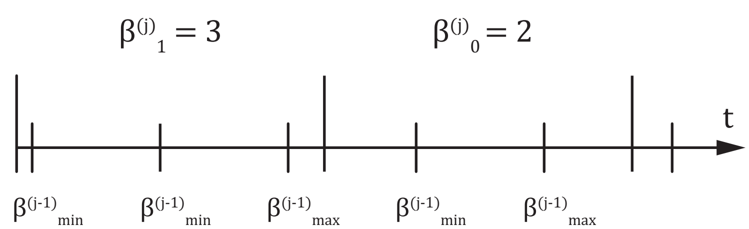

FIGURE 12. A general $\beta^{(j)}$ sequence.

6. SATISFIABILITY CONDITIONS

In this paper, we will only present a satisfiability condition for collision sequences, where α starts and ends with v, and all $\beta^{(j)}$ start and end with $\beta^{(j)}_{max}$. This is only a minor restriction, and applying our results to the general case would only require some special treatment for the ends of the collision sequence.

Theorem 6.1. *Given a collision sequence that starts and ends with v, and all $\beta^{(j)}$ start and end with $\beta^{(j)}_{max}$, the collision sequence is valid $\leftrightarrow$ the following is true for all j*

$$\beta^{(j)}_i \in \left\{ \left\lfloor \frac{a_{j-2}}{a_{j-1}} \right\rfloor, \left\lceil \frac{a_{j-2}}{a_{j-1}} \right\rceil \right\} \tag{14}$$

Proof. Proving the $\rightarrow$ part of the theorem follows directly from Lemma 4.1.

To prove the $\leftarrow$ part we can notice that α is valid if and only if $\delta^{(0)}$ is valid, and $\delta^{(0)}$ is valid if and only if $\delta^{(1)}$ is valid, etc. Going back to Lemma 4.1 again, $\delta^{(j)}$ is valid only if Equation 14 is true. □

Theorem 6.2. *For every valid collision sequence,* $\lim_{n\to\infty} a_n = 0$

Proof. From the definition of a_j

$$a_j = \beta^{(j-2)}_{max}\, a_{j-1} - a_{j-2} \tag{15}$$

$$= \left\lceil \frac{a_{j-2}}{a_{j-1}} \right\rceil a_{j-1} - a_{j-2} \tag{16}$$

Now from the definition of the ceiling function, we know that

$$0 \leq \left\lceil \frac{a_{j-2}}{a_{j-1}} \right\rceil - \frac{a_{j-2}}{a_{j-1}} < 1 \tag{17}$$

Rearranging, we get

$$0 \leq \left\lceil \frac{a_{j-2}}{a_{j-1}} \right\rceil a_{j-1} - a_{j-2} < a_{j-1} \tag{18}$$

Combining the above equation with Equation 15, we get

$$0 \leq a_j < a_{j-1} \tag{19}$$

Thus a is strictly decreasing and bounded below by 0, so $\lim_{n\to\infty} a_n = 0$. □

7. Continued Fractions

Finally, we will show how collision sequences relate to continued fractions. A simple continued fraction of a real number r (hereafter referred to as a continued fraction) is the expression given below in equation 20:

$$r = k_1 + \frac{1}{k_2 + \frac{1}{k_3 + \cdots}} \tag{20}$$

Where $k_1 = \lfloor r \rfloor, r_1 = \frac{1}{r-k_1}, k_2 = \lfloor r_1 \rfloor, r_2 = \frac{1}{r_1 - k_2}, \ldots$. This process terminates when r_i becomes an integer. The integers k_i are called "partial quotients" and a real number r can be expressed in its continued fraction form $r = [k_1, k_2, k_3, \ldots]$. As an example, we can find the continued fraction for $r = 3.245$.

Example 7.1.

$$3.245 = 3 + \frac{1}{4 + \frac{1}{12 + \frac{1}{4}}} \tag{21}$$

In this example, we see that $k_1 = \lfloor r \rfloor = \lfloor 3.245 \rfloor = 3$ *by the definition. Thus, we can define* $r_1 = \frac{1}{3.245-3} = \frac{1}{0.245} = \frac{200}{49}$ *and find* $k_2 = \lfloor r_1 \rfloor = \left\lfloor \frac{200}{49} \right\rfloor = 4$. *Now we can recursively define* $r_2 = \frac{1}{\frac{200}{49} - 4} = \frac{49}{4}$. *As we continue onwards in this process, we find* $k_3 = 12$, $r_3 = 4$, *and* $k_4 = 4$. *Thus, the continued fraction representation is* $3.245 = [3, 4, 12, 4]$.

If we use the continued fraction representation for the slope m of a combined trajectory for a billiard ball, then we have:

$$m = \lfloor m \rfloor + \frac{1}{\lfloor 1/\{m\} \rfloor + \frac{1}{\lfloor 1/\{1/\{m\}\} \rfloor + \cdots}} \tag{22}$$

Where the partial quotients are given by:

$$k_1 = \lfloor m \rfloor$$
$$k_2 = \left\lfloor \frac{1}{\{m\}} \right\rfloor$$
$$k_3 = \left\lfloor \frac{1}{\left\{\frac{1}{\{m\}}\right\}} \right\rfloor$$
$$k_4 = \left\lfloor \frac{1}{\left\{\frac{1}{\left\{\frac{1}{\{m\}}\right\}}\right\}} \right\rfloor$$
$$\vdots$$

We see that the partial quotients are given by the recursive formula: $k_j = \left\lfloor \frac{1}{m-[k_1,k_2,\ldots,k_{j-1}]} \right\rfloor$. In fact, a more interesting observation is that the sequence of partial quotients k_j form exactly the sequence of the minimum number of tick marks in each $\beta^{(j)}$ subproblem:

Theorem 7.2. *Given the continued fraction representation $m = [k_1, k_2, k_3, \ldots]$ for the slope $m \in \mathrm{R}$ of the combined trajectory T of a billiard ball, we must have $k_j = \beta^{(j)}_{min}$.*

Proof. We will proceed by induction. This is clearly true for $k_1 = \lfloor m \rfloor$ by lemma 4.1, since for $\beta^{(j)}_{min}$ the spacing of the larger tick marks are m and the spacing of the smaller tick marks is 1.

Now suppose we have shown our hypothesis to be true for all $k_1, k_2, \ldots, k_{j-1}$. This means that $\beta^{(i)}_{min} = k_i$ for all $i \leq j-1$. By definition 5.3 (of the sequence $\beta^{(j)}$), we see that $\beta^{(j)}_i$ is one more than the number of occurrences of $\beta^{(j-1)}_{max}$ between the ith and $(i+1)$st occurrence of $\beta^{(j-1)}_{min}$ in the $\beta^{(j-1)}$ sequence.

Now, we can find an expression for k_j based on our definition of continued fractions:

$$k_j = \left\lfloor \frac{1}{m - [k_1, k_2, \ldots, k_{j-1}]} \right\rfloor \tag{23}$$
$$= \left\lfloor \frac{1}{m - [\beta^{(1)}_{min}, \beta^{(2)}_{min}, \ldots, \beta^{(j-1)}_{min}]} \right\rfloor \tag{24}$$

However, notice that $d = m - [\beta^{(1)}_{min}, \ldots, \beta^{(j-1)}_{min}]$ is equal to the spacing between tick marks in the $\beta^{(j)}$ sequence. Thus, we see that $\beta^{(j)}_{min} =$

$\lfloor 1/d \rfloor$. However, this is exactly what we wanted to show, since we know that $k_j = \lfloor 1/d \rfloor$, so we see that $\beta^{(j)}_{min} = k_j$. □

In fact, one can see that the process of finding $\beta^{(j)}_i$ for the $\beta^{(j)}$ subproblems exactly mirrors the process of finding the partial quotients k_j in a continued fraction of m.

8. Conclusion

In this paper we explored properties of valid billiard ball collision sequences. We introduced a useful tiling representation that drastically simplified the problem and allowed us to easily characterize all valid collision sequences. From the tiling representation, we developed a series of "meta" problems in the same manner as one calculates a continued fraction expansion. We also showed an interesting way to visualize the original problem and the "meta" problems as 1-dimensional problems on a number line.

References

[1] Continued Fraction. (2013, 11 30). Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Continued_fraction