SEQUENCES OF BILLIARD BALL COLLISIONS

JONATHAN ALLEN, JOHN WANG

Abstract. In this paper we explore properties of sequences of billiard ball collisions. We present a tiling representation which is used to help elucidate some simple properties of these sequences. Then, we provide a one-dimensional representation which builds upon the tiling representation. Next, we provide a method which takes an arbitrary sequence and checks to see if it could have been created by a billiard ball colliding with a billiard table. Finally, we show how these sequences relate to continued fractions.

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.

Date: December 6, 2013.

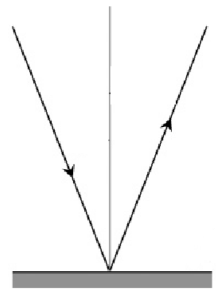

FIGURE 1. Mechanics of a Billiard Ball Colliding with a Table Edge

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:

Definition 1.1. *A* collision sequence *(α) for a ball B is the sequence of v's and h's which appear as the ball collides with the walls of the billiard table.*

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 (for example when $\mathbf{v}_0 = (1, 0)$) 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 $\mathbf{v}_0 = (0.23, 0.05)$.

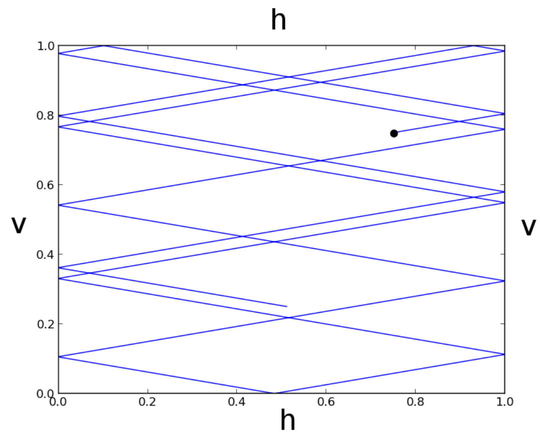

FIGURE 2. Example Billiard Ball Trajectory, $x_0 = (0.75, 0.75)$, $v_0 = (0.23, 0.05)$

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:

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

A valid collision sequence given these initial conditions would start and end with an h collision. A possible example would be:

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

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

$$\alpha = hvvvvvhvvvvvhvvvvvh \tag{3}$$

1.3. **Simplifications.** There are some simplifications that we can make which 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. Thus, we can characterize $\mathbf{v}_0$ by the angle γ that $\mathbf{v}_0$ makes with the positive horizontal line (just as in polar coordinates). From now on, we will use $\mathbf{v}_0$ and γ interchangeably.

- 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 $\pi + \gamma \in [0, \pi/2]$ (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. For the rest of the paper, we will assume $\gamma \in [0, \pi/2]$.

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.

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 unit square billiard table on the xy plane ($[0, 1]^2$). 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

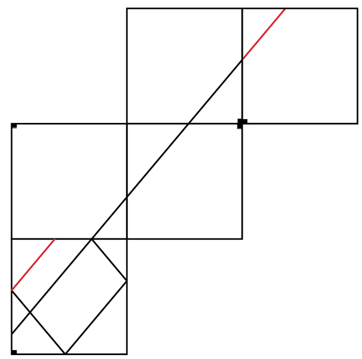

FIGURE 3. Example of Tiling Billiard Tables

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

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

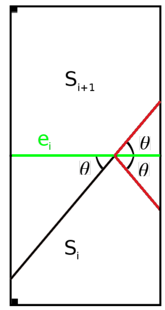

FIGURE 4. Trajectories $\tau(\kappa_{i-1}, \kappa_i)$ in s_i and $\tau(\kappa_i, \kappa_{i+1}$ s_{i+1} forming a line segment

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 $i \in \mathrm{Z}$.

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 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.

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 $\kappa^v_k = t_{v,k}$, where κ^v_k 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, $\kappa^h_k = t_{h,k}$.

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.

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 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:

$$k = \begin{cases} q & \text{if } p \pmod 2 \equiv 0 \\ 2q & \text{if } p \pmod 2 \equiv 1 \end{cases} \tag{4}$$

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 both 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 $\lfloor (m(x+1)+y_0) - (mx+y_0) \rfloor = \lfloor m \rfloor \geq 1$. 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. □

We therefore see that if v occurs consecutively in a collision 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 v 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 v (i.e. sequences like *hvvvhvvvh* become *vhhhvhhhv*) without loss of generality.

4. 1-Dimensional Representation

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

$$x(t) = \frac{1}{m}\,t + x_0 \tag{5}$$

$$y(t) = t \tag{6}$$

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

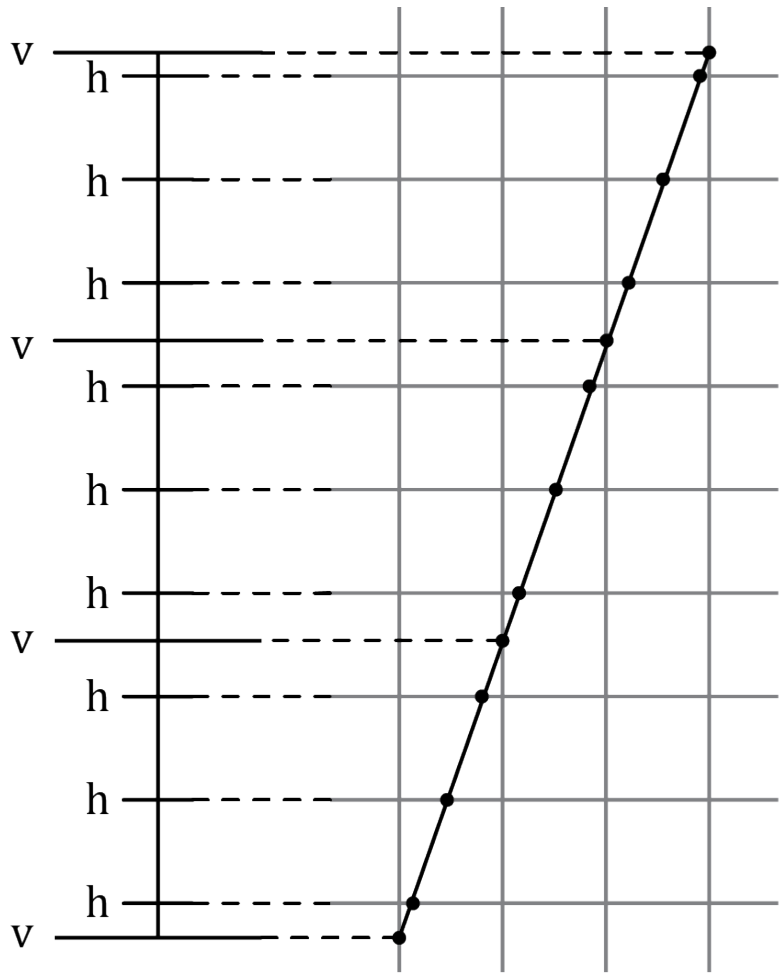

FIGURE 5. Projecting onto the parametric representation.

For the sake of space, we will rotate the t axis so that it is horizontal.

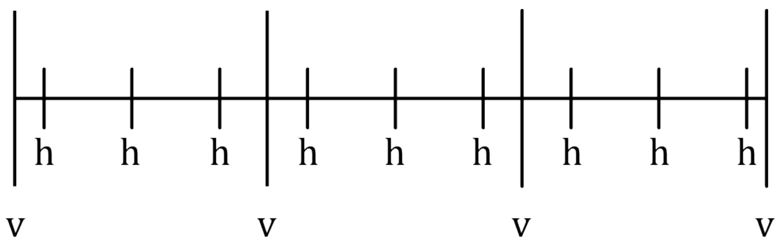

FIGURE 6. Projecting onto the parametric representation.

Lemma 4.1. *Let $l_2 < l_1$. The number of real numbers spaced at regular intervals l_2 inside an 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 spaced at regular intervals l_2 inside an interval of length l_1. Given some $a \in A$, let $b \in \mathbb{R}$ be defined such that the following holds

$$l_1 = (a-1)l_2 + b \tag{7}$$

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

Rearranging Equation 7, we get

$$a = \frac{l_1}{l_2} + 1 - \frac{b}{l_2} \tag{8}$$

Which leads directly to

$$a \in \left\{\left\lfloor\frac{l_1}{l_2}\right\rfloor, \left\lceil\frac{l_1}{l_2}\right\rceil\right\} \tag{9}$$

□

Lemma 4.2. *A finite 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.*

Proof. TODO: can extend any collision sequence to any length. □

Because of Lemma 4.2, without loss of generality we can confine ourselves to only look at collision sequences that start and end with an h.

Definition 4.3. *Given a collision sequence α, define a sequence $\beta^{(0)}$ where each element $\beta_i^{(0)}$ is the number of h collisions between the i^{th} and $(i+1)^{th}$ v in α.*

The $\beta^{(0)}$ sequence is much simpler to think of geometrically. Looking at Figure 7, $\beta_i^{(0)}$ represents the number of h tick marks in between each v tick mark.

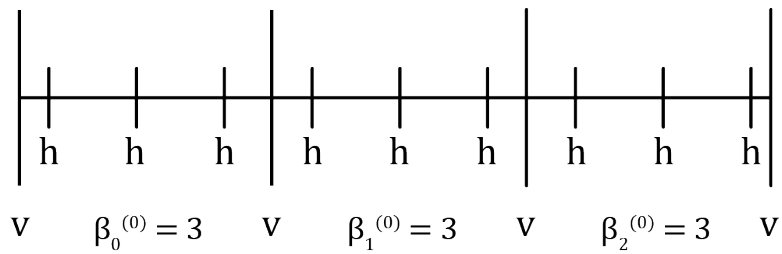

FIGURE 7. The $\beta^{(0)}$ sequence.

Definition 4.4. *Given a valid collision sequence, define the sequence $\beta^{(j)}$ where each element $\beta_i^{(j)}$ is 1 more than the number of occurrences of $\beta_{max}^{(j-1)}$ between the i^{th} and $(i+1)^{th}$ occurrence of $\beta_{min}^{(j-1)}$ in the $\beta^{(j-1)}$ sequence.*

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

An example $\beta^{(j)}$ sequence is plotted in Figure 8.

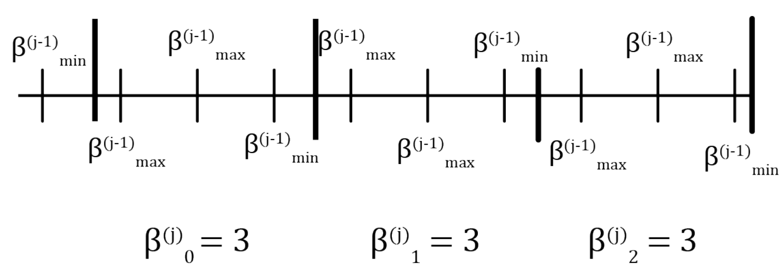

FIGURE 8. The $\beta^{(j)}$ sequence.

Now we can derive a very useful sequence that we will call a. We start with $a_0 := m, a_1 := 1$. These two values represent the two intervals sizes in our parametric representation.

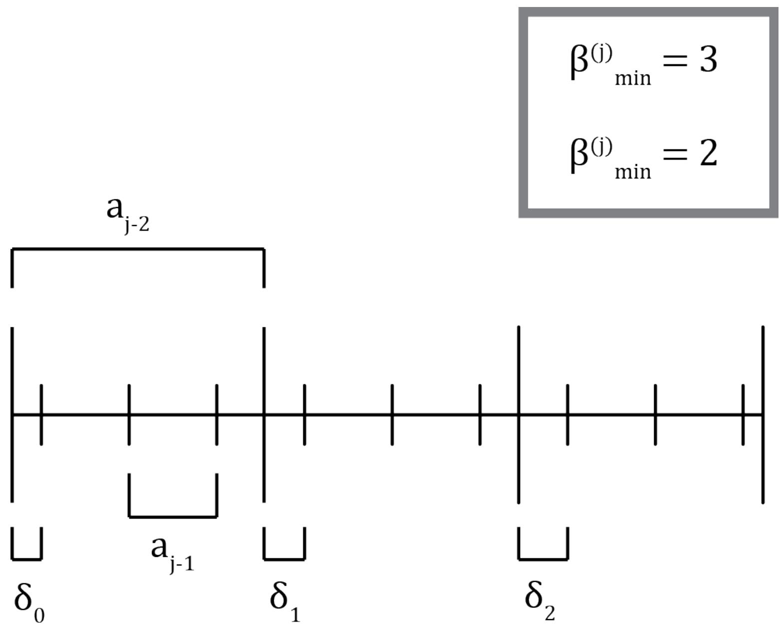

FIGURE 9. The $\beta^{(i)}$ sequence.

From Lemma 4.1, we know that $\beta^{(0)}_{max} = \lceil \frac{m}{1} \rceil$.

$$\delta_i^{(0)} := \begin{cases} x_0 & \text{for} \quad i = 0 \\ i(\beta^{(0)}_{max} * 1 - m) + x_0 & \text{for} \quad i \geq 1 \end{cases} \tag{10}$$

We can notice that

$$\beta_i^{(0)} = \left\lfloor \delta_i^{(0)} \right\rfloor + \beta^{(0)}_{max} - \left\lfloor \delta_{i+1}^{(0)} \right\rfloor \tag{11}$$

Thus if we plot the values of $\delta^{(0)}$ we can see an interesting pattern of the β

The rest of a will be defined inductively. For $2 \leq j < j_f$, assume that a_{j-1}, a_{j-2}.

$$\delta_i^{(j)} := \begin{cases} x_0 & \text{for} \quad i = 0 \\ i(\beta^{(j)}_{max} * 1 - m) + x_0 & \text{for} \quad i \geq 1 \end{cases} \tag{12}$$

The number of tick marks in each interval is $\beta^{(0)}$ Now from 4.1, we know that

$$\gamma = \left\lceil \frac{a_{j-1}}{a_{j-2}} \right\rceil \tag{13}$$

γ represents the smallest integer number of intervals of length a_{j-1} that can cover an interval of length a_{j-2}. Thus at most γ tick marks with spacing a_{j-1} can fit inside an interval of length a_{j-2}.

This sequence is written concisely in the following definition

Definition 4.5. *Define the sequence a*

$$a_j := \begin{cases} m & \text{for} \quad j = 0 \\ 1 & \text{for} \quad j = 1 \\ \beta_{max}^{(j-2)} a_{j-1} - a_{j-2} & \text{for} \quad 2 \leq j < j_f \end{cases} \tag{14}$$

From now on we will only consider collision sequences, where each non-terminal $\beta^{(j)}$ starts and ends with $\beta_{min}^{(j)}$.

Theorem 4.6. *A collision sequence is valid iff the following is true for all j*

$$\beta_i^{(j)} \in \left\{ \left\lfloor \frac{a_{j-2}}{a_{j-1}} \right\rfloor, \left\lceil \frac{a_{j-2}}{a_{j-1}} \right\rceil \right\} \tag{15}$$

Proof. This was already proved above, we are just stating it again. □

Theorem 4.7. *For every valid collision sequence,* $\lim_{n\to\infty} a = 0$

Proof. From the definition of a_j

$$a_j = \beta_{max}^{(j-2)} a_{j-1} - a_{j-2} \tag{16}$$

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

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

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

Rearranging, we get

$$0 \leq \left\lceil \frac{a_{j-2}}{a_{j-1}} \right\rceil a_{j-1} - a_{j-2} < a_{j-1} \tag{19}$$

Combining the above equation with Equation 16, we get

$$0 \le a_j < a_{j-1} \tag{20}$$

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

5. Continued Fractions

Finally, we will show how collision sequences relate to continued fractions.

6. Conclusion

TODO