# UNIFORMITY OF CUBE LINES AND RELATED PROBLEMS

**BY MICHAEL SAMUEL DONDERS**

**A dissertation submitted to the**

**Graduate School—New Brunswick**

**Rutgers, The State University of New Jersey**

**in partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**Graduate Program in Mathematics**

**Written under the direction of**

**József Beck**

**and approved by**

_____

_____

_____

_____

**New Brunswick, New Jersey**

**May, 2017**

# ABSTRACT OF THE DISSERTATION

Uniformity of cube lines and related problems

By Michael Donders

Dissertation Director: József Beck

We define a cube line to be a geodesic traveling over the surface of the cube; that is—we take a straight line traveling on one of the faces of a cube, and when it hits an edge it continues on to the next face so that if the two faces were unfolded to be coplanar, the two line segments on either face would connect to form a straight line. The principal question we look to answer is: does this cube line uniformly distribute over the surface of the cube. Here, we define uniformly distributed as, for any Jordan measurable test set, the proportion of the cube line which lies in the test set approaches the relative size of the test set as the length of the cube line approaches infinity. This problem was derived as an extension to the classical problems of uniform distribution of a torus line over a unit square and uniform distribution of billiard paths over a unit square. The arguments in this problem, however, are quite different from these previous problems, and take ideas from many fields, including erogidc theory, number theory, geometry and combinatorics.

# ACKNOWLEDGEMENTS

I would foremost like to thank my advisor, József Beck, with whom this work was jointly done, and whose help and assistance was invaluable to me in writing this dissertation. I would like to thank my mathematics professors at Rutgers University and at McDaniel College for teaching me mathematics. And I would like to thank Andrew Donders for his time and help in editing this paper.

# Contents

# CHAPTER 1

# INTRODUCTION

We study a typical geodesic on the surface of a cube; it is a curve which consists of straight line segments on each of the six faces of the cube, such that when switching from one face of the cube to another, the two line segments on either side of the edge shared between two neighboring faces meet at that edge and do so in a basically collinear fashion. This is to say, if the two neighboring faces were to be made coplanar by rotation about that shared edge, the two line segments on either face would form a single line segment. We call this geodesic a "cube surface line", or simply a "cube line". See Figure 1.1 for an example of an initial segment of a cube line.

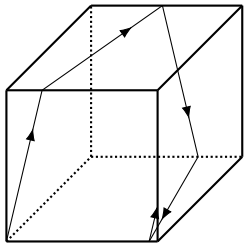

Figure 1.1: Initial segment of a cube line traveling around a cube.

Much work has been done on the properties of a straight torus line on the two-dimensional unit torus $[0,1)^2$, see Figure 1.2; and the cube being a three dimensional structure comprised of six squares, it is natural to see the cube line as an analog to a torus line in a square. The uniformity of a torus line on a square is well understood, thanks principally to a breakthrough result by H. Weyl in 1916, where in his famous paper he characterized the Weyl Criterion, a

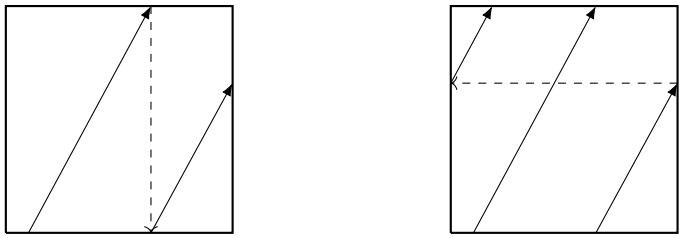

Figure 1.2: A torus line on the unit square. When the line hits the top edge, it translates down to the bottom edge, and when it hits the right edge it translates across to the left edge.

general criterion that implies equidistribution, or uniform distribution, which, when applied to the problem of torus lines on a square, implies that all straight torus lines with irrational slope are uniformly distributed over a square [20]. To be more precise, the result states that for any polygonal test set $\Omega$ in the unit square torus $[0,1)^2$, and any torus line with irrational slope $\alpha$ and arbitrary starting location, the proportion of the line in the test set is equal to the area of the test set (area refers to the two-dimensional Lebesgue measure). Constant speed parametrization of the line by the function $f(t)$ allows us to formally state this as,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f_\Omega(t)dt = |\Omega|,$$

where,

$$f_\Omega(t) = \begin{cases} 1 & f(t) \in \Omega \\ \\ 0 & f(t) \notin \Omega \end{cases}.$$

The Weyl Criterion proves this equidistribution claim, as well as the generalized claim for arbitrary dimensions; and, moreover, beyond this qualitative result of equidistribution, has a quantitative result as well, providing bound speed of convergence to equidistribution, i.e. upper bounds are given to corresponding error terms. In spite of the large amount of work on the uniformity of a torus line in a square and billiard paths in a square, the similar problem of a straight line on the surface of a cube has remained stubbornly open for the last 100 years.

In this paper, we prove a series of results which yield increasingly powerful results about the cube line similar to those of the torus line on the square provided by Weyl. We prove: there exists a cube line which is dense on the surface of the cube; there exists a cube line which is uniformly distributed on the surface of the cube; all cube lines are dense over the surface of the

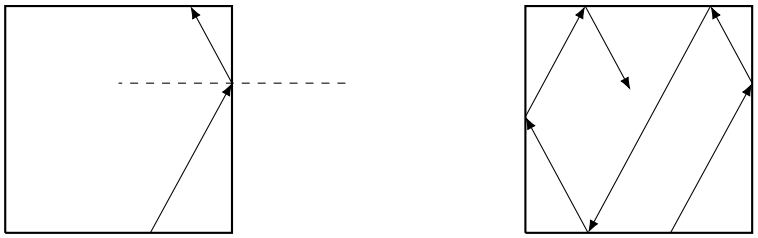

Figure 1.3: A billiard path the unit square. When the path hits an edge, it reflects off of it with the angle of reflection equal to the angle of incidence

cube; all cube lines which have irrational slopes are uniformly distributed on cube symmetric sets (to be defined later) on the surface of the cube; all cube lines that have irrational slopes and bounded continued fraction digits are uniformly distributed; all cube lines that have irrational slopes are uniformly distributed—the central result of the paper; and finally several generalizations of the result. Additionally, in Chapters 5 and 6 we include a look at some quantitative results for a specific class of cube lines. The quantitative results are particularly interesting because the application of ergodicity via Birkhoff's ergodic theorem gives only "soft" qualitative results. Our work uses different ideas from a large variety of topics, lying at the cross-roads of uniform distribution, geometry, number theory and combinatorics.

Our main results are: Theorems 1-2 (see page 12), Theorems 3 (see page 38), Theorems 4 (see page 46), Theorems 5 (see page 70), Theorems 6 (see page 70), Theorems 7-8 (see page 84), Theorems 10 (see page 118), Theorems 11 (see page 120), Theorems 12 (see page 121). The proofs of Theorems 1-2 and Theorems 5-6 in particular are very lengthy and involved.

We briefly review some known results about billiard paths on the unit square, a well studied and understood problem thanks principally to results by Weyl, König and A. Szücs [20][14]. These results serves as both the foundation of the motivation of this paper, and will also be a useful theorem used throughout.

The concept of a 2-dimensional "billiard path" more precisely describes a point-mass (which represents a tiny billiard ball) moving freely along in a straight line inside the unit square with unit speed until it hits one of the edges of the square. This collision with the boundary

is elastic, meaning it follows the laws of reflection—angle of incidence equals angle of reflection. After the reflection, the point-billiard continues on in a straight line with its new velocity, maintaining its constant speed from before, until it hits another boundary, and then this process repeats, see Figure 1.3. (We consider these billiards in ideal circumstances and ignore all types of dampening, e.g. friction, air resistance.) The initial conditions of the billiard, the starting location $s$ and the initial direction $\theta$, uniquely determine an infinite piecewise linear billiard path $x(t)$ in the unit square. The law of reflection implies that there are at most four possible directions the billiard can travel in—the initial direction is preserved modulo $\pi/2$ during any reflection. Because of the unit speed of the movement, $t$ represents both time and arc-length.

Formally, a billiard path within a unit square $[0,1]^2$ has the form

$$x(t) = (x_1(t), x_2(t)), 0 < t < \infty \text{ with } x_j(t) = 2||(s_j + t\beta_j)/2||, \quad j = 1,2,$$

where $\mathbf{e} = (\beta_1, \beta_2)$ is a unit vector, and $||y||$ denotes the distance of a real number $y$ from the nearest integer. In this form, $\mathbf{s} = (s_1, s_2)$ is the starting point and $\mathbf{e}$ is the initial direction. Thus $\theta = \arctan \frac{\beta_1}{\beta_2}$ is the initial angle.

An alternative more symmetric way to represent this is to replace $[0,1]^2$ with $[-\frac{1}{2}, \frac{1}{2}]^2$. In this case we have

$$x(t) = (x_1(t), x_2(t)), 0 < t < \infty \text{ with } x_j(t) = < s_j + t\beta_j >, \quad j = 1,2,$$

where

$$< y >= ||y|| \text{ if } y \geq 0, \text{ and } -||-y|| \text{ if } y < 0.$$

We next examine the geometric trick of unfolding a billiard path inside the unit square into a straight line on the entire plane. The idea is simple: when the billiard line hits an edge, rather than reflecting the line, we reflect the square itself along the respective edge. This procedure results in a given billiard path being unfolded into a straight line in the plane, see Figure 1.4.

Two straight lines in the plane correspond to the same billiard path if and only if they differ by a translation through an integral vector where both coordinates are even, i.e. where the vector is from the double square lattice $\{2\mathbb{Z}\}^2$. In this way we can reduce the problem of

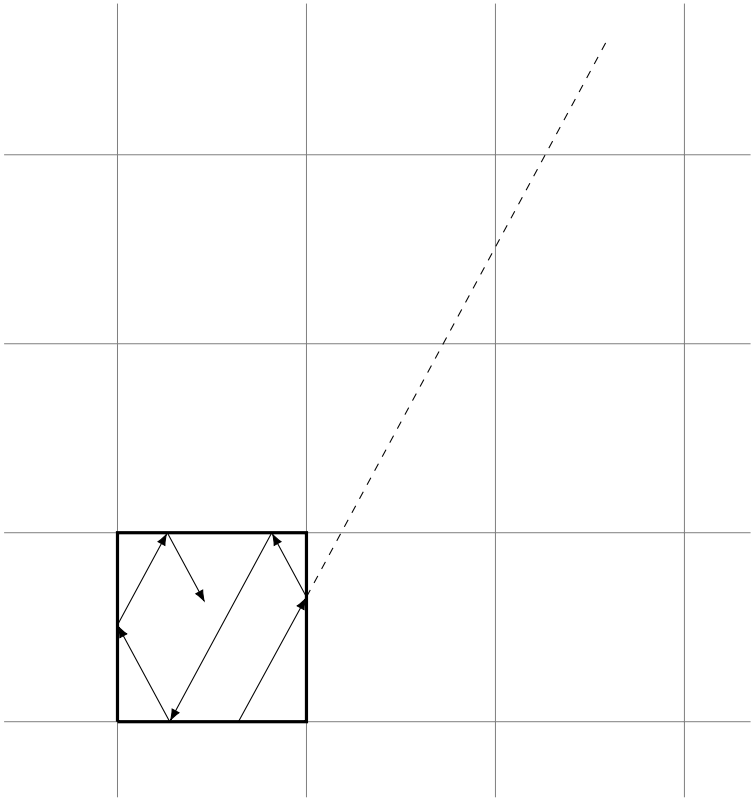

Figure 1.4: Unfolding the unit square onto the unit grid turns the billiard path into a straight line.

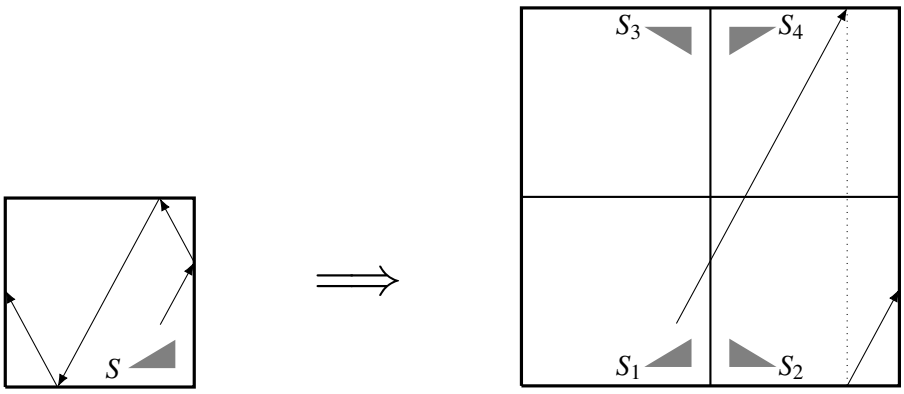


Figure 1.5: The grid is equivalent to every 2×2 interval, which means the billiard path on the unit square is equivalent to a torus line on a 2×2-square, with each test set in the unit square reflected into four congruent test sets on the 2×2-square.

distribution of a billiard path in the unit square to the distribution of the corresponding torus line in the $2 \times 2$ square.

The problem of uniformity of a billiard path in the unit square with respect to a given test set $S$ is equivalent to the uniformity of the torus line in a $2 \times 2$ square with respect to four set tests, $S_1, S_2, S_3, S_4$, which are copies of the test set $S$ reflected along each of the edges, or rather across each of the integer lines of the $2 \times 2$ square, see Figure 1.5. The $2 \times 2$ square is then shrunk (scaled linearly) into the unit square $[0,1)^2$.

In the general case, the test set is upgraded to a function $f(x,y) \in L([0,1)^2)$, which is a continuous or periodic real-valued Lebesgue integrable function with period 1 in both variables. In our case, $f$ is the $0-1$ valued characteristic function of the union $S_1 \cup S_2 \cup S_3 \cup S_4$, of the four shrunk and reflected copies of the test subset $S$.

To generalize to higher dimensions, note that a billiard path in the unit cube $[0,1]^d$, $d \geq 3$, can be defined similarly. We can unfold the line into a straight line in the $d$-dimensional plane by reflecting the cube across the $d-1$-dimensional edge rather than reflecting the line itself. Again, in this way we can equate uniformity in a test set $S$ of the $d$-dimensional unit billiard path to uniformity of the torus line in the $d$-dimensional 2 x 2 x ... x 2 hypercube, where each of the $2^d$ copies of the unit cube contains a reflected copy of the test set $S$.

As far as we know, the first appearance of the geometric trick of unfolding is in a paper of D. König and A. Szücs from 1913 [14], and became widely known after Hardy and Wright included it in their well-known book on number theory [10]. The continuous generalization of

the classical equidistribution theorem (Kronecker-Weyl) implies an elegant property of the torus line in the unit square: if the slope of the initial direction is rational, then the path is periodic, but if the initial slope is irrational, then the torus line is dense (Kronecker's Theorem), and, what is more, for irrational slope the torus line is also uniformly distributed over the unit square [20]. Precisely, this means that for any axis-parallel rectangular test set $S = [a,b] \times [c,d] \subset [0,1]^2$,

$$\lim_{T\to\infty} \frac{1}{T} \text{measure}\{t \in [0,T] : x(t) \in S\} = \text{area}(S) = (b-a)(d-c),$$

where we have $x(t),\ 0 < t < \infty$ denoting the torus line in the unit square,
parametrized by its arc-length (=time), $t$. This upgrade of Kronecker's Theorem to include uniformity, due to H. Weyl, represents the classical definition of uniform distribution, is called Weyl type uniformity (introduced in Weyl's classical paper in 1916)[20].

König and Szücs used the trick of unfolding, combined with the Kronecker-Weyl theorem, to prove the following analogous property for the billiard path in the unit square: if the initial direction of the path is rational, then the path is periodic, but if the initial slope is irrational, then the torus line is uniformly distributed (and thus also dense) in the unit square. This completes the 2-dimensional case and this argument generalizes to work in every dimension $d \geq 3$.

One particular reason this problems has remained unsolved is, despite its similarities to the torus line and billiard paths there is one large difference: the difference between removable singularities and proper singularities. Consider a torus line on a square which hits a vertex of the square. By applying a small translation to the torus line in either direction, we can eliminate this issue, resulting in a uniquely defined extension of the line. In this sense the vertex is a 'removable singularity', only potentially causing ambiguity if the line actually hits the singularity, but importantly still allowing translation across the singularity to be continuous (using any reasonable metric to compare distance between torus lines, such as maximum distance or average distance).

In contrast we observe the case of a cube line hitting a vertex of the cube. If we try to alleviate the issue in a similar manner, we find that when applying a small translation we are faced with ambiguity in how to do so. A small translation up and a small translation down will result

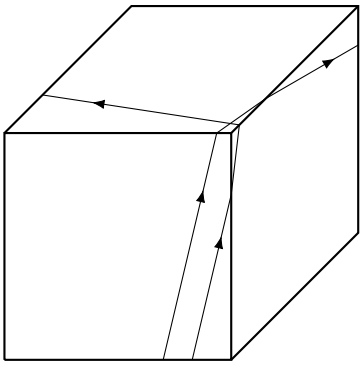

Figure 1.6: Each corner of the cube represents a singularity. Here we can see how two cube lines which are initially very close translations of each other but are split by a singularity become drastically different.

in vastly different lines, see Figure 1.6. We can in fact see that translation across a vertex of a cube is not a continuous mapping (again using any reasonable metric on the distance between two cube lines, again such as maximum distance or average distance). In this sense the vertices of a cube are 'proper singularities', in that they are not only issues when the line actually hits the singularity, but also anytime a line is translated across a singularity.

The underlying reason for this difference between the square and the surface of the cube is that the plane can be tiled with the square but not with the surface of the cube. We can see this further in the case of a line on the surface of a regular tetrahedron. An unfolded tetrahedron is an equilateral triangle, which can tile the plane, and therefore the vertices of the tetrahedron are 'removable singularities', in that translation over these singularities is again a continuous mapping. Because this technique of unfolding is available, the problem of a geodesic on the surface of the tetrahedron has been solved by Beck [2], which also includes several generalizations of the problem.

Moreover, how close the cube line is to a vertex is an equivalent question to how close a multiple of its slope (plus a possible fixed constant) is to an integer. The fact that irrational rotations mod 1 (with a fixed added constant) are uniformly distributed over [0,1) indicates that all cube lines with irrational slope will get arbitrarily close to a vertex of the cube, and thus any translation of the cube line will yield a non-continuous result which is exceptionally hard to track. This is the essential issue which makes this a challenging problem, as being unable to translate makes knowing where the cube line is at any given time extremely difficult.

Again comparing the two cases, supposing we have a cube line parametrized by $L(t)$, and a line on a square torus parametrized by $\Lambda(t)$, both with the same unit speed and same irrational slope $\alpha$. If $\Lambda(0) = 0 \times \beta \in \{0\} \times [0,1)$, then we know that $\Lambda(t) = (t \mod 1) \times (B + \alpha n \mod 1)$, giving us the exact location of the line at any point. However, if $L(0) = \gamma \in [0,1)$ along some edge $[0,1)$ of the cube, we cannot say exactly where $L(t)$ is. We can generally say where it will be on a face—it will be at $(t \mod 1) \times (B + \alpha n \mod 1)$, but we know neither which face it is on, nor what the orientation of the face is. At best we can say it is one of 48 points which together form a cube symmetric set—that is, the set of possible locations of the geodesic is invariant under action by the symmetric group of the cube. We will discuss cube symmetric sets further in Chapter 8.

We note that in 1984 E. Gutkin [8] announced a positive solution to the problem of a geodesic on a cube, but, unfortunately, the key step in his paper remains unproved: its proof was not published ever since (email communication by W. Veech; we will return to Gutkin's paper later—see the end of Chapter 10 and the end of Chapter 11). Now 33 years later it is time to clarify this situation, especially since Gutkin unfortunately died in 2013, Veech died in 2016, and Gutkin's paper has been quoted in several survey papers as an important work which gives the "complete solution of the almost integrable case." As far as we know, our paper is the first complete proof of the uniformity of cube lines (as well as the first with quantitaive results, and generalizations far beyond the cube–see Chapters 5, 6 and 11).

We begin with a formal statement of the cube line, and the notation used in this paper.

In the usual way, we refer to the six faces of the unit cube $[0,1]^3$ by calling the face with $x = 1$ the front face (or just front), $x = 0$ the back face, $z = 1$ the top face, $z = 0$ the bottom face, $y = 1$ the right face, and finally $y = 0$ the left face.

Assume that the starting point, $s$, for a geodesic, $L$, is $L(0) = \mathbf{s} = (1, s_2, s_3)$, $0 < s_2, s_3 < 1$, and the initial direction of the geodesic is described by the unit vector $\mathbf{v} = (v_1, v_2)$. Note that an initial point and direction uniquely parametrize all geodesics on the cube surface. For ease of reference, we refer to the 'slope' of $L$ as $\beta = v_2/v_1$, such that $(1, \beta)$ is parallel to $\mathbf{v}$, or by

angle $\theta$, such that $\tan(\theta) = 1/\beta$. The parametrization of $L$, for small $t$, looks as follows:

$$L(t) = (1, s_2 + v_1 t, s_3 + v_2 t) \tag{1.1}$$

and this holds so long as $0 < s_2 + v_1 t, s_3 + v_2 t < 1$.

By symmetry we may assume that $v_1 > 0$ and $v_2 > 0$, thus we first hit an edge at $\min\{(1-s_2)/v_1, (1-s_3)/v_2\}$, hitting the top edge and moving to the top face in the case of the minimum being $t = (1-s_2)/v_1$, and hitting the right edge and moving to the right face in the case of $t = (1-s_3)/v_2$ being the minimum. This motion of the cube generates an infinite "quasi-random alternating" sequence of the six faces. As said before, this sequence is very difficult to track, resulting in it being very difficult to find which face the cube line is on at any given point (and thus the location of the cube line on the cube) without brute force checking of the exact location of the cube line by looking at every face the cube line enters, and which face it enters next, one-by-one. This issue is discussed further in Chapter 3.

In general, we assume the geodesic is parametrized with unit constant speed (starting from **s**), and thus the parameter $t$ represents both time and distance, with $L(t)$ denoting the location of the geodesic at time $t$.

To see how the parametrization looks after the geodesic moves from one face to another, assume $(1-s_3)/v_2 > (1-s_2)/v_1$. Then the geodesic will move from the front face to the right face. Thus on the next face, $y = 1$. We note that the $z$ value does not actually change, and remains $z = (1-s_2)/v_1$; finally the $x$ value starts at $x = 1$ and decreases at a rate of $v_3$. We get the formula

$$x = 1 - v_1\left(t - \frac{1-s_2}{v_1}\right) = 1 - v_1 t + 1 - s_2 = -(s_2 + v_1 t),$$

which holds until the geodesic hits its next edge at $z = 1$ or $x = 0$. We see that there are always three different types of coordinates: a constant coordinate, a $v_1$-variable coordinate and a $v_2$-variable coordinate, based upon what the coefficient of $t$ is in each coordinate. The cube line hits an edge when one of the variable coordinates equals 0 or 1, where the coordinate swaps its variable coefficient with the constant coordinate. In the above example, on the front face with direction $(v_1, v_2)$, $x$ was the constant coordinate, $y$ the $v_1$ coordinate and $z$ the $v_2$. Upon $y = 1$, $y$ and $x$ swapped so on the right face $y$ was the constant coordinate and $x$ was the $v_1$ coordinate.

While it becomes very difficult to track this parametrization due to the quasi-random alternation of faces generated by $L(t)$ discussed above we note that there exists a limited number of directions in which the line can be traveling on each face. On, say, the front face, the directional vector will always be one of the following four alternatives: vector $\mathbf{v} = (v_1, v_2)$, the perpendicular vector $\mathbf{v}^\perp = (v_2, -v_1)$, and their negatives $-\mathbf{v} = (-v_1, -v_2)$, $-\mathbf{v}^\perp = (-v_2, v_1)$. These four vectors represent four half-lines, and only two different line-directions, $\pm\mathbf{v}$ and the perpendicular $\pm\mathbf{v}^\perp$. Of course this symmetrically holds for the other faces as well.

For a given vector $\mathbf{v}$, this defines the $(\pm\mathbf{v}, \pm\mathbf{v}^\perp)$-geodesic flow, or simply the $\mathbf{v}-$flow, on the surface of the cube.

We quickly mention two negligible cases we will ignore. Firstly, when $\alpha$ is rational, it becomes very clear that the cube line is simply periodic, and thus neither dense nor uniformly distributed. This follows directly from the fact that the rotation $r + \alpha n \mod 1$ is periodic for rational $\alpha$. In this case the possible locations where the geodesic can hit the edges is finite, and there are a limited number of directions with which it can hit each point, so the geodesic must be periodic as well (due to the Markov-like property of the line at these locations, which follows from a point and a direction uniquely defining a cube line). As there are only countably many rational slopes, this however represents a negligible set.

The second case we disregard is when the cube line $L(t)$ hits a vertex of the cube at some point $t > 0$; as discussed previously, this results in a proper singularity that cannot be uniquely resolved. This will occur when $(s_2, s_3)$ and $(v_1, v_2)$ are rationally dependent. We call such cases *pathological* and, for any fixed irrational slop, the set of starting locations which are rationally dependent is again a countable set, and thus is again negligible. We will, however, allow the initial point of a cube line to be a vertex, since as long as the initial direction and which face it is initially traveling on is well defined, there is no ambiguity. Moreover, since irrational rotations (i.e. rotations of an irrational number on the unit interval [0,1)) beginning at an integer can never hit an integer again, this implies any cube line with irrational slope beginning at a vertex will indeed never hit another vertex.

We end the chapter with the formal statement of the theorem we look to prove for Weyl type uniformity of the cube line, as well as an analogous result we will also prove regarding Birkhoff type uniformity.

**Theorem 1** (Weyl Type Uniformity Result)**.** *If $\alpha = v_2/v_1$ is irrational, for almost every starting point $L(0)$, the relative time the geodesic $L(t)$ spends in a given Jordan measurable test set as $0 \leq t \leq N$, $N \to \infty$ tends to the limit density in case of uniform distribution on the cube surface, i.e. it tends to the relative area of the test set.*

**Theorem 2** (Birkhoff Type Uniformity Result)**.** *If $\alpha = v_2/v_1$ is irrational, for almost every starting point $L(0)$, the relative time the geodesic $L(t)$ spends in a given Lebesgue measurable test set as $0 \leq t \leq N$, $N \to \infty$ tends to the limit density in case of uniform distribution on the cube surface, i.e. it tends to the relative 2-dimensional Lebesgue measure of the test set.*

The slight difference between these theorems being Theorem 1 is restricted to Jordan measurable test sets; in contrast, Theorem 2 allows for Lebesgue measurable test sets. Moreover, both theorems are limited to almost every starting point. It is clear that for a given starting point, there are some Lebesgue measurable test sets for which relative time the geodesic spends in it does not tend to the measure of the set. For example if we take the set of the surface of cube minus the cube line: the line itself has 2-dimensional Lebesgue measure zero so this test set would have full Lebesgue measure (six), but the geodesic would also, by construction, spend zero time within the test set. Note though that Theorem 2 says that for any given Lebesgue measurable test set, for all but a measure zero set of starting locations of the geodesic, the time it spends in the test set will asymptotically be equal to the relative measure of the test set. In the example test set above, for a given slope the test set would fail for one particular geodesic (based on initial point), but the set of starting points for which a geodesic of that slope would fail on is measure zero.

We therefore have that Theorem 2 must allow this measure zero exception. This measure zero set of exceptions for any given test set includes both pathological starting locations, as well as other potentially problematic starting locations, as in the given example. In Theorem 1, on the other hand, it is not immediately clear that this requirement is necessary, as the example above is not Jordan measurable. In fact in general, uniformity in the Weyl sense would include all points. While Theorem 1 does include the measure zero set exception, as we will see in Theorem 9, this can indeed be strengthened to include all non-pathological starting points, which

still must be held as an exception because of their undefined nature.

For some far-reaching generalizations of Theorems 1-2, see Chapter 11.

# CHAPTER 2

# DISCRETIZATION

One useful tool we will use is taking this problem of looking at a continuous geodesic over the 2-dimensional cube surface and translating it to the discrete problem of looking at where the geodesic hits on the 1-dimensional set of edges of the cube. For a geodesic with initial angle $\theta$, we can assume by symmetry that $0 < \theta < \pi/2$. There are two ways the geodesic can hit an edge of the cube: with angle $\theta$, or with angle $\theta^{\perp} = \theta + \pi/2$ (angle here refers to the angle between the line segment and the edge). If the crossing has angle $\theta$ we call it a **$\theta$-edge crossings**, and if it has angle $\theta^{\perp}$ we call it a **$\theta^{\perp}$-edge crossings**, see Figure 2.1. These two crossings correspond to the two potential line pairs of slope vectors $\pm\mathbf{v}$ and $\pm\mathbf{v}^{\perp}$.

We further note that for each edge, there are two ways to cross the edge at a given angle. For example, on the edge shared by the front and right faces, a geodesic can cross at angle $\theta$ from the front face to the right face, or it can cross at angle $\theta$ from the right face to the front face. This corresponds to the two possible directional vectors $\mathbf{v}$ and $-\mathbf{v}$. We call a pair $\vec{\mathscr{E}} = (\mathscr{E}, \mathbf{v})$ of an edge and a slope a **directed $\theta$ edge** , or **directed $\theta^{\perp}$ edge**, depending on the angle and direction of the slope. This gives us 24 total directed $\theta$ edges, two for each edge of the cube. Similarly there are 24 total directed $\theta^{\perp}$ edges as well. Lastly, we sometimes wish to talk about all crossings on an edge in a particular direction regardless of the angle. In this case we will refer to a $\vec{\mathscr{E}}$ as simply a **directed edge**, meaning we are looking at all $\theta$- and $\theta^{\perp}$-edge crossings that occur on the given edge in one particular direction.

We will initially look at only the $\theta$-edge crossings and ignore the $\theta^{\perp}$-edge crossings, though we will get symmetrical results for $\theta^{\perp}$-edge crossings. One fact we can note about $\theta$-edge

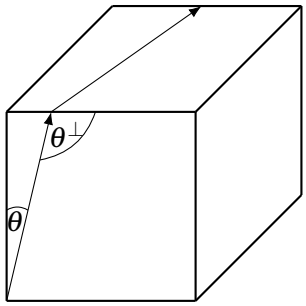


Figure 2.1: There are two ways the cube line can cross an edge of cube: with angle $\theta$, or with angle $\theta^\perp$.

crossings is that, if there is a $\theta$-edge crossing from face $A$ to face $B$, then the next $\theta$-edge crossing will be from some face $C$ (possibly, but not necessarily, the same face as $B$) into the face opposite of $A$. For example, suppose there is a $\theta$-edge crossing going from the bottom face into the front face. We know that the next time the geodesic hits an edge at an angle $\theta$ will be when it hits one of the top edges to go into the top face, though we do not know from which face that transition occurs.

Another way of looking at the differences between $\theta$-edge crossings and $\theta^\perp$-edge crossings is, for a geodesic with starting location $\mathbf{s} = (1, s_2, s_3)$, and initial direction $(v_1, v_2)$ (again with $0 < \theta = \arctan(v_2/v_1) < \pi/2$), a $\theta$-edge crossing occurs at $L(t)$ whenever $s_3 + v_2 t = 0 \mod 1$, while $\theta^\perp$-edge crossings occur whenever $s_2 + v_1 t = 0 \mod 1$. To see this, it is clear that an edge crossing occurs if and only if $s_2 + v_1 t = 0 \mod 1$ or $s_3 + v_2 t = 0 \mod 1$. If $s_3 + v_2 t = 0 \mod 1$, then the angle between the edge and the line is $\arctan(v_2/v_1) = \theta$, while if $s_2 + v_1 t = 0 \mod 1$, then the angle between the edge and the line is $\arctan(v_1/v_2) = \theta^\perp$.

From this perspective we can show the above claim that a $\theta$-edge crossing from face $A$ to face $B$ will next go into the face opposite $A$. Suppose by symmetry a $\theta$-edge crossing goes from the bottom face to the front face. This must occur when $s_3 + v_2 t = 0$. As seen in the parametrization in 1.1, at an edge-crossing the constant coordinate swaps with the variable coordinate. This means we start with $z = 0$ being the constant (i.e. the bottom face), which then swaps with the $v_2$ coordinate $x$, which follows since the $\theta$-edge crossings occur at $s_3 + v_2 t = 0$ on the edge between the bottom and front faces. Thus $z$ becomes the $v_2$ coordinate and $x$ becomes the constant coordinate. Therefore the next time $s_3 + v_2 t = 0 \mod 1$ will be when $s_3 + v_2 t = 1 = z$; thus the next $\theta$-edge crossing will occur on the constant edge $z = 1$, i.e. the

top edge which is opposite of the bottom edge.

By only looking at the $\theta$-edge crossings , we can track the movement of the geodesic starting at a $\theta$-edge crossing by unfolding the faces the geodesic can travel on before it hits the next $\theta$-edge crossing into a cycle. If, again for example, there is a $\theta$-edge crossing from the bottom face to the front face, we know the next $\theta$-edge crossing will go into the top face. Thus between the two $\theta$-edge crossings the cube line travels through the front, right, back and left faces. These together form a what we call a **cycle**, a $4 \times 1$ loop with the $\theta$-edge crossings occurring where the line hits on the bottom and top of this loop. The top and bottom of this cycle are comprised each of four directed $\theta$ edges . Labeling each directed $\theta$ edge as the half open interval $[0,1)$ in the obvious way, it becomes clear that if one $\theta$-edge crossing occurs at $\beta \in [0,1)$ in some directed $\theta$ edge, the next $\theta$-edge crossing occurs at $\beta + 1/\alpha \mod 1$ in some other directed $\theta$ edge. From this we can see that by looking only at where on each directed $\theta$ edge each $\theta$-edge crossing occurs, we get the sequence $(\beta + n\alpha \mod 1)$, where $\beta$ is some constant depending on the initial point of the cube line. We will see this technique of unfolding a cycle of four faces more in Chapter 3.

This sequence of $\theta$-edge crossings represents an arithmetic progression in $t$ such that the time gap between each of the progressions is

$$\frac{1}{\sin(\theta)} = \sqrt{1+\alpha^2}.$$

We wish to reduce the problem of proving that the cube line is uniformly distributed over the surface of the cube to the discrete problem of proving that the sequence of $\theta$-edge crossings is uniformly distributed over the edges of the cube (or more specifically, the directed $\theta$ edges of the cube).

Assume a particle moves with constant speed on a geodesic and enters face **F** of the cube through edge $\mathscr{E}$ at angle $\theta$. By symmetry, we can for simplicity assume that $\mathscr{E}$ and geodesic are such that the line is crossing from the bottom face into the front face, **F**, and the angle $\theta$ satisfies $\pi/4 < \theta < \pi/2$. The positive orientation of **F** induces a direction onto the edge $\mathscr{E}$, giving us directed $\theta$ edge $\vec{\mathscr{E}}$, and the segment of geodesic on **F** gives a directed line segment.

We identify the directed $\theta$ edge $\vec{\mathscr{E}}$ with the unit interval $[0,1]$ in the standard way, using the orientation of the directed edge to determine which end is 0, and do similarly for all 24 other directed $\theta$ edges .

We now suppose there was a $\theta$-edge crossing at any given point on this interval $\vec{\mathscr{E}}[0,1]$ and see where the next $\theta$-edge crossing would be. A (theoretical) $\theta$-edge crossing that occurred at the left endpoint, 0, of the interval of $\vec{\mathscr{E}}$ would next hit $\alpha = 1/\tan(\theta) \mod 1$ on the unit interval corresponding to the directed $\theta$ edge on which the next $\theta$-edge crossing lies; call this directed $\theta$ edge $\vec{\mathscr{E}}_1$. As we assumed $\pi/4 < \theta < \pi/2$ we can see that $0 < \alpha = 1/\tan(\theta) < 1$, and thus this will simply hit the point $\alpha$ on the edge directly above $\mathscr{E}$ in the process of going from the front face to the top face. Similarly, a (again theoretical) $\theta$-edge crossing occurring at the right endpoint, 1, of the interval, would next hit the point $1-\alpha$ on the unit interval corresponding to to directed $\theta$ edge on which the next $\theta$-edge crossing lines; call this directed $\theta$ edge $\vec{\mathscr{E}}_2$, which as $0 < \alpha < 1$ will be the directed $\theta$ edge from the right face to the top face. Note the directed $\theta$ edge of the next $\theta$-edge crossing for the left and right endpoint, $\vec{\mathscr{E}}_1$ and $\vec{\mathscr{E}}_2$, will be different but adjacent, sharing a single point. We can note that the point $1-\alpha$ on $\vec{\mathscr{E}}$ would next hit the right endpoint of $\vec{\mathscr{E}}_1$, and also hits the left endpoint of $\vec{\mathscr{E}}_2$ (i.e. the shared point of these edges). Other than this one shared point, we can see that this process induces a linear mapping on $\vec{\mathscr{E}}$, mapping $(0,1-\alpha)$ in $\vec{\mathscr{E}}$ to $(\alpha,1)$ in $\vec{\mathscr{E}}_1$, and mapping $(1-\alpha,1)$ in $\vec{\mathscr{E}}$ to $(0,\alpha)$. This mapping is essentially an $\alpha$-shift, $x \to x+\alpha \mod 1$, see Figure 2.2. Formally

$$\vec{\mathscr{E}}(0,1-\alpha) \to \vec{\mathscr{E}}_1(\alpha,1), \quad \vec{\mathscr{E}}(1-\alpha,1) \to \vec{\mathscr{E}}_2(0,\alpha). \tag{2.1}$$

We extend this mapping to all 24 directed $\theta$ edges, resulting in a piecewise unit linear mapping from the set of directed $\theta$ edges to itself, comprised of 24 disjointed $\alpha$-shifts. We glue the unit intervals of the 24 directed $\theta$ edges together in the obvious way to form the interval [0,24); applying an arbitrary ordering to them and having,

$$\vec{\mathscr{E}} = \vec{\mathscr{E}}_0 = [0,1),\ \vec{\mathscr{E}}_1 = [1,2),\ \vec{\mathscr{E}}_2 = [2,3),\ \vec{\mathscr{E}}_3 = [3,4),\ \ldots,\ \vec{\mathscr{E}}_i = [i,i+1),\ \ldots, \tag{2.2}$$

for integer $0 \le i < 24$. Thus the 24 directed $\theta$ edges together form the interval $[0,24)$ and the mapping above, with the slight alteration into half-open intervals, defines an interval exchange

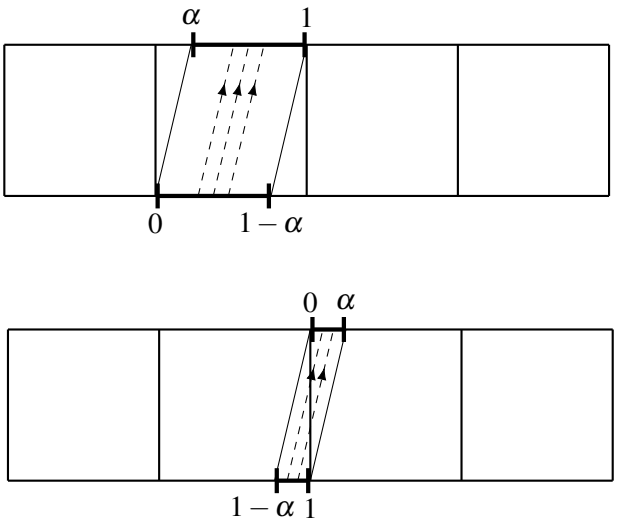


Figure 2.2: The mapping $T$, which takes one $\theta$-edge crossing to the next can be decomposed into piecewise-linear measure-preserving shifts. The sub-interval $(0, 1-\alpha)$ on an edge gets translated to $(\alpha, 1)$, and $(1-\alpha, 1)$ gets translated to $(0, \alpha)$. This specific example assumes $\alpha < 1$, but similar logic follows for all values of $\alpha$.

transformation

$$T : [0, 24) \to [0, 24). \tag{2.3}$$

More precisely, $T$ translates a half open interval of the form $[r-1, r-\alpha)$, for $r$ an integer, $1 \leq r \leq 24$, to some other half-open interval of the form $[r'-1+\alpha, r')$, $r'$ again an integer with $1 \leq r' \leq 24$, and translates half-open intervals of the form $[r-\alpha, r)$ to half-open intervals of the form $[r''-1, r''-1+\alpha)$, once again with $r''$ an integer, $1 \leq r'' \leq 24$. The mapping $T$ has discontinuous jumps at the points $x$, where the fractional part of $x$, $\{x\} = 1-\alpha$ or, $\{x\} = 0$. We call these jumps the singular points of $T$. We emphasize the crucial fact that $T$ is a Lebesgue measure preserving transformation.

The key characteristic of the transformation $T$ is the slope of the geodesic which induces the irrational shift, $0 < \alpha < 1$, so we denote the shift $T = T_\alpha$. Note this shift can similarly be done on any irrational slope, not only $0 < \alpha < 1$, which was just a result of our simplifying assumption that, using symmetry, we can use take $\pi/4 < \theta < \pi/2$. Notably, a transformation can similarly be constructed on the $\theta^\perp$-edge crossings of the geodesic; we denote this transformation $\widetilde{T} = \widetilde{T}_{-1/\alpha}$.

We now claim we can reduce the problem of proving Weyl Type Uniformity on the surface

of the cube (i.e. Theorem 1) to proving the transformation $T$ (or equally $\widetilde{T}$) is ergodic.

**Lemma 1.** *For a cube line with slope $\alpha$, if the transformation $T_\alpha$ is ergodic over the interval $[0,24)$, then the cube line is uniformly distributed in the Weyl sense (see Theorem 1).*

As we already know that $T$ is measure preserving, $T$ being ergodic implies that for any point $x \in [0,24)$, the sequence of images of the point $(T^n(x))_{n=0}^{\infty}$ is uniformly distributed (formal proof of this standard result is given at the end of Chapter 4). Or, for any starting point $x \in [0,24)$, and any interval $I \subseteq [0,24)$,

$$\lim_{N\to\infty} \frac{|\{T^n(x) : 0 \le n \le N\} \cap I|}{n} \to \frac{|I|}{24}.$$

Consider as a test set a tilted rectangle $\mathscr{A}$ on a face $\mathbf{F}$ of the cube such that, for $\vec{\mathscr{E}}$ a directed $\theta$ edge on the face $\mathbf{F}$, the rectangle $\mathscr{A}$ has two parallel sides with angle $\theta$ to $\vec{\mathscr{E}}$, and two parallel sides with angle $\theta$ to one of the directed $\theta$ edges of $\mathbf{F}$ perpendicular to $\vec{\mathscr{E}}$ (i.e. angle $\theta + \pi/2$ to $\vec{\mathscr{E}}$). See Figure 2.3.

A straightforward geometric consideration shows that the amount of relative time the cube line spends in the test set $\mathscr{A}$ from directed $\theta$ edge $\vec{\mathscr{E}}$ is equal to the relative time the sequence $x, Tx, T^2x, T^3x, \ldots$ spends in a sub-interval $I'$ of $\vec{\mathscr{E}} \subseteq [0,24)$, the length of which will be equal to the length of $\mathscr{A}$ (note here from its construction as a tilted rectangle, the width of $\mathscr{A}$ does not matter as width will scale equally with both relative time the geodesic is in it and size of $\mathscr{A}$). As this sequence is uniformly distributed, and thus relative to the size of the sub-interval $I'$, this implies the relative time the geodesic is in $\mathscr{A}$ from $\vec{\mathscr{E}}$ is equal to the relative size of $\mathscr{A}$. As all Jordan measurable sets can be constructed through these tilted rectangles, this generalizes to all Jordan measurable test sets.

There are a few caveats worth mentioning here: first, Birkhoff's theorem is about a single (but arbitrarily complicated Lebesgue measurable) test set, while Weyl's uniformity is about all simple sets. But every interval can be approximated by rational intervals, of which there are only countably many, so applying Birkhoff's ergodic theorem for $T$ with rational intervals, the union of the measure zero sets with exceptions (i.e. bad starting locations) is still a zero set. Thus Birkhoff's theorem still implies Weyl uniformity. See end of Chapter 3 for more detail on the resolution to this caveat.

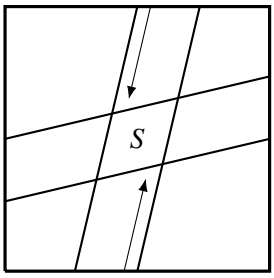


Figure 2.3: Uniform distribution of edge crossings on the directed edges implies uniform distribution of the cube line over tilted rectangles on the surface of the cube. As all polygonal test sets can be well approximated by these tilted rectangles, this also implies uniform distribution over all test sets on the surface of the cube.

Secondly, it should be noted that $T$ being uniformly distributed actually implies a slightly stronger result than the cube line being uniformly distributed. The above argument only looks at one way in which the geodesic enters the test set $\mathscr{A}$, from $\vec{e}$. But there are other ways the geodesic can enter $\mathscr{A}$ as well, for example, from the directed $\theta$ edge opposite $\vec{e}$ on face $\mathbf{F}$ pointed in the other direction. What this lemma implies is that for any test set, the relative time the geodesic spends in the test set from each directed $\theta$ edge is proportional to the size of the interval on the edge which leads to $\mathscr{A}$. Summing these contributions over all edges implies Weyl uniformity over the cube surface, but the original statement is a slightly stronger result as the inverse implication does not immediately follow. □

We prove a similar lemma for Birkhoff uniformity (Theorem 2).

**Lemma 2.** *For a cube line with slope $\alpha$, if the transformation $T_\alpha$ is ergodic over the interval $[0,24)$, then the cube line is uniformly distributed in the Birkhoff sense (see Theorem 2).*

To prove Birkhoff type uniformity, it suffices to show that the geodesic flow, for a given irrational slope $\alpha$ over the entire surface of the cube is ergodic—that is, showing that any subset $\mathbf{S}^*$ of the surface of the cube which is invariant under the geodesic flow must have full measure or measure zero.

For a given irrational slope $\alpha$, suppose the geodesic flow is not ergodic. Then there exist a set $\mathbf{S}^*$ which is invariant under the geodesic flow with $0 < \lambda^2(\mathbf{S}^*) < 6$, where $\lambda^2$ refers to the 2-dimensional Lebesgue measure.

Let $S \subset \mathbf{S}^*$ be the set of directed $\theta$-edge crossings restricted to the flow in the invariant set

$\mathbf{S}^*$. It is easy to see $S$ is measurable and has non-trivial measure—that is, $0 < \lambda(S) < 24$ where $\lambda$ is the 1-dimensional Lebesgue measure, as a consequence of $\mathbf{S}^*$ having non-trivial measure; if $S$ has measure zero, then $\mathbf{S}^*$ must have measure zero, and similarly if $S^c$, the complement of $S$, has measure zero, then $(\mathbf{S}^*)^c$ will have measure zero as well. But by the construction of $S$, $S$ must be invariant under $T$, and thus neither $S$ nor $S^c$ has measure zero. This shows that $T$ being ergodic sufficiently implies geodesic flow over the surface of the cube is also ergodic, and thus Birkhoff uniformly distributed. □

# CHAPTER 3

# META-LINES

In this chapter we introduce an interesting relationship between two cube lines, and introduce a "special" cube line $A(t)$ with starting point a vertex of the cube and initial slope $\alpha = -2+\sqrt{5}$ (why this cube line in particular is special will be explained later).

We first formally define a **cycle** of the cube, as briefly mentioned in Chapter 2, to be any four faces of the cube which form a loop, with an orientation but no specific starting point. For example, front-right-back-left faces form a cycle, and would be the same cycle as right-back-left-front, but a different cycle from front-left-back-right. As an example, Figure 1.1 in Chapter 1 shows a cube line traveling through the front-top-back-bottom cycle from the left end of the cycle to the right end. We note the cube contains 6 unique cycles. If we think of the cube line as traveling on these cycles, then they are interesting because we can define the $\theta$-edge crossings of a cube line to be precisely when the cube line leaves one cycle and enters a new cycle. We also note that at any given point, the cube line is traveling through two cycles. For example if the cube line is on the front face, moving in the positive $y$ and $z$ directions (to the upper right), then it is in both the front-right-back-left cycle, and the front-top-back-bottom cycle. These two cycles are distinctly different, though, distinguished by the angle at which the cube line entered the cycle: in one it enters at angle $\theta$, and the other at angle $\theta^{\perp}$. We will assume the cycle the cube line is traveling in is the one entered at angle $\theta$.

The four faces of a cycle of the cube together act as a $1 \times 4$ torus with respect to the cube line, as mentioned in Chapter 1. For example, if $L(t)$ hits a $\theta$-edge crossing on the directed edge going from the bottom face to the front face, before it hits its next $\theta$-edge crossing it will

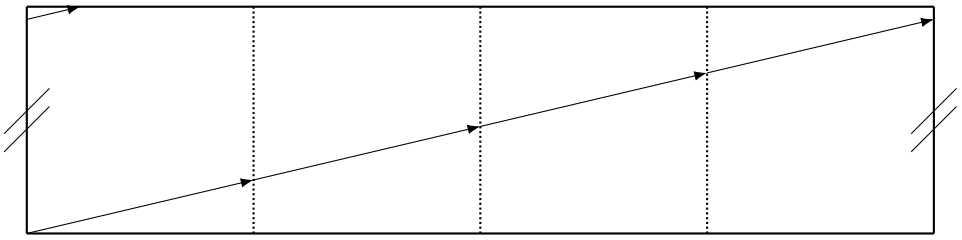

Figure 3.1: The segment of a cube line traveling along a single cycle can be thought of as a line traveling on a 4x1 torus.

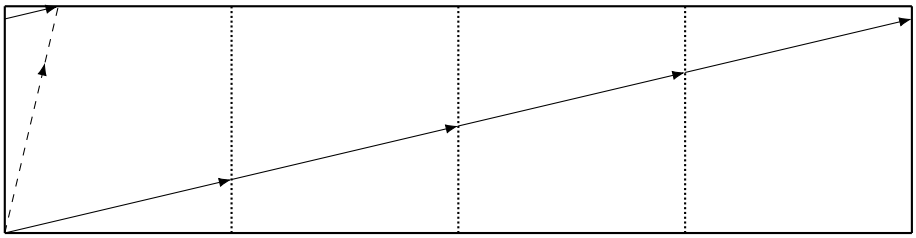

Figure 3.2: Tracking the relationship between the original line and its meta-line can be done by observing their behavior on the 4x1 torus cycles.

travel through the front, right, back and left faces, until it hits any edge leading into the top face, which will be the next $\theta$-edge crossing. These four faces together make a cycle where, if the line hits the right end of the cycle, it acts like a torus line and goes back to the same point on the left edge. This interpretation of the cycles of the cube makes it very easy to see where consecutive $\theta$-edge crossings will be, see Figure 3.1.

We define a new concept we call a **meta-line** for a particular set of cube lines: cube lines whose initial point $L(0) = V$ is a vertex, and whose initial angle $\alpha$ is either greater than 4 (in absolute value) or less than 1/4 (in absolute value). Suppose a cube line $L(t)$ with slope $\alpha$ satisfies these conditions. We define the meta-line of $L$ as follows: if $|\alpha| > 4$ (by symmetry assume $\alpha > 4$), we assume that $L(0) = V$ is the bottom left corner of the front face and the initial direction of $L(t)$ is in the upper-right direction (this can also be assumed by symmetry). We define the meta-line of $L$ to be the cube line $L^*(t)$ with slope $\alpha^* \equiv \alpha$ modulo 4, and initial point $L^*(0) = V^* = V$ (meta-line definition when $|\alpha| < 1/4$ will follow shortly). The key thing to note about this meta-line is that

$$L\left(n\sqrt{1+\left(\frac{1}{\alpha}\right)^2}\right) = L^+\left(n\sqrt{1+\left(\frac{1}{\alpha^*}\right)^2}\right) \text{ for all } n \in \mathbb{N}; \tag{3.1}$$

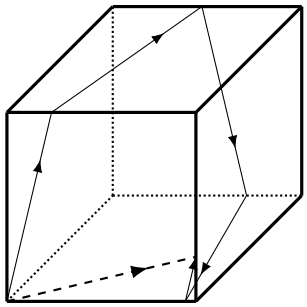

Figure 3.3: A cube line traveling on a cube and its meta-line. A cycle of the cube line corresponds to a single face crossing of the meta-line.

that is, $L^*$ hits all of the $\theta$-edge crossings of $L$. This can be shown inductively, noting that $L(0) = L^*(0)$ trivially: between the points $L(n/\alpha)$ and $L((n+1)/\alpha)$ for each $n \in \mathbb{N}$, considering the cycle the line travels on being a $1 \times 4$ torus, it is clear that $L$ and $L^*$ will hit the same location at their next $\theta$-edge crossing since a $1 \times 4$ torus will ignore a difference modulo 4, see Figure 3.2. The idea behind this definition of a meta-line is to remove extra loops through the cycles, while maintaining the same $\theta$-edge crossings of the cube line.

Alternatively, if $|\alpha| < 1/4$, again we assume $\alpha > 0$ and assume that $L(0) = V$ is on the bottom left vertex of the front face. We define the meta-line of $L$ to be $L^+(t)$, the cube line with slope

$$\alpha^+ = \frac{1}{\frac{1}{\alpha} \text{ modulo } 4},$$

and initial point $L^+(0) = V^+ = V$. This is essentially the definition of the meta-line for $|\alpha| > 4$ if we were to rotate and mirror the cube so that $L$ goes from a line with $|\alpha| > 4$ to one with $|\alpha| < 1/4$. From this interpretation we can see that

$$L\left(n\sqrt{1+\alpha^2}\right) = L^+\left(n\sqrt{1+(\alpha^+)^2}\right) \text{ for all } n \in \mathbb{N}. \tag{3.2}$$

This can again be seen when looking at the cycles the line travels over as a $4 \times 1$ or $1 \times 4$ torus, and we see again this definition of a meta-line is removing a cycle from this torus, thus leaving us with the same $\theta$-edge crossings.

See Figures 3.3 and 3.4 for examples of deriving meta-lines from cube lines. Some notes on meta-lines: firstly, as the meta-line of $L(t)$ will hit all the $\theta$-edge crossings of $L(t)$, it follows from the results in Chapter 2 that the meta-line of $L(t)$ is uniformly distributed if $\theta$-edge

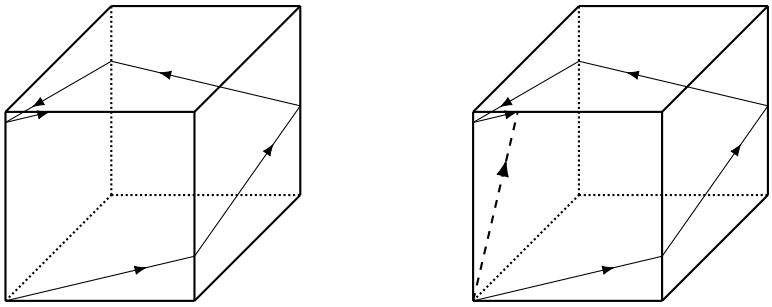

Figure 3.4: The initial segment of a cube line with initial slope $|\alpha| < 1/4$ and its meta-line.

crossings of $L(t)$ are uniformly distributed, and $L(t)$ is uniformly distributed if the angled edge crossings of its meta-line are uniformly distributed (note they are no longer $\theta$-edge crossings as the angle of the meta-line is different from that of the original).

Secondly, they are called meta-lines because they tell us where the line will be "faster" than the line will itself. For example, if $\alpha > 4$, as in 3.1, we know $L\left(n\sqrt{1+\left(\frac{1}{\alpha}\right)^2}\right)$ and $L^*\left(n\sqrt{1+\left(\frac{1}{\alpha^*}\right)^2}\right)$ will be in the same spot. But $|\alpha| > 4$ and $\alpha^* \equiv \alpha \mod 4$ implies that $n\sqrt{1+\left(\frac{1}{\alpha}\right)^2} > n\sqrt{1+\left(\frac{1}{\alpha^*}\right)^2}$. Thus the length of $L$ will be much larger than the length of $L^*$, despite the fact that they will be in the same spot. Another way to look at this is to say a cube line crossing through an entire cycle of four faces any number of times is equivalent to its meta-line crossing just a single face (i.e. eliminating extra loops through a cycle). A similar comparison of 3.2 shows the same thing: the original lines and the meta-line will be in the same spot while the meta-line will be much shorter than the original. Again, we see that the comparison is–an entire cycle in the original line is reduced to a single face crossing in the meta-line.

Thus, if we want to find what points a cube line will hit, it is quicker to look at its meta-line (if applicable). But further, in some cases once we take the meta-line, we can indeed take the meta-line again (e.g. an irrational number slightly more than 4), which would still maintain the $\theta$-edge crossings of the original line, and in fact tell us where they are even quicker. Each time we iterate this process the new meta-line provides information even more quickly in comparison to the original line. This motivates us to look for a line where we can repeatedly and indefinitely find its meta-line.

It would be helpful to find a line on which the meta-line process is recursive. First we can

ask: is there any line which is its own meta-line? The answer to this is clearly no, as, if $\alpha > 4$, then $\alpha \mod 4$ cannot also be greater than 4, with similar logic following for $\alpha < 1/4$. The next question to ask is: is there a line which is the meta-line of its meta-line? To this question we find the answer is actually, yes.

Let $L(t)$ be a cube line with starting point $L(0) = V$ being on the front face in the bottom-left corner, and $L$ have irrational slope $\alpha$ with $|\alpha| > 4$. We want to know if there is a value for $\alpha$ such that $L = L^{*+}$. As the slope of $L$ is $\alpha$, the slope of $L^*$ will be $\alpha \mod 4$. Then the slope of $L^{*+}$ will be

$$\frac{1}{\frac{1}{\alpha \mod 4} \mod 4}.$$

Setting this equal to the slope of $L$ we get,

$$\alpha = \frac{1}{\frac{1}{\alpha \mod 4} \mod 4}.$$

There are in fact many solutions to this, found by taking $\alpha \mod 4 = \alpha - 4k_1$, and similarly, $\frac{1}{\alpha - 4k_1} \mod 4 = \frac{1}{\alpha - 4k_1} - 4k_2$, for $k_1, k_2 \in \mathbb{Z}/\{0\}$. Then we get

$$\alpha = \frac{1}{\frac{1}{\alpha - 4k_1} - 4k_2}$$

$$\Rightarrow \frac{1}{\alpha} = \frac{1}{\alpha - 4k_1} - 4k_2.$$

Solving for $\alpha$ yields

$$\alpha = \frac{2k_1k_2 - \sqrt{4k_1^2k_2^2 + k_1k_2}}{k_2}. \tag{3.3}$$

This represents a family of solutions that will work; each values $k_1$ and $k_2$ correspond to how many loops through a cycle are removed by the meta-line process. Moreover, this family is the solution which are the meta-lines of their meta-lines, but longer recurrence chains are possible, too, e.g. $L = L^{*+*+}$, viz. any line which is its own fourth meta-line (in fact any even length chain is possible). The proceeding arguments about meta-lines work for any geodesic in this family of solutions, but for ease of notation we chose a single solution to work with, using $k_1 = k_2 = -1$ in 3.3, giving us

$$\alpha = \sqrt{5} - 2. \tag{3.4}$$

Let $A = A(t)$ be the particular cube line with slope $\alpha$ and initial point a (fixed) vertex $V$. From the definition in 3.2, its meta-line is a cube line with the same starting point and slope $\beta = 2 + \sqrt{5}$; call this cube line $B = B(t)$. Note that $\beta \equiv \alpha$ modulo 4 and $\beta = 1/\alpha$. We now have two lines, $A$ and $B$, where $B$ is the meta-line of $A$, and $A$ is the meta-line of $B$. For the rest of the paper, cube lines $A = A(t)$ and $B = B(t)$ will refer to these specific cube lines. (As an aside, for clarity all figures in this paper are drawn with the cube lines having slope $\alpha$ or $\beta$).

Assume that $\theta$ is the angle for cube line $A$, then $\phi = \theta^{\perp}$ is the angle for cube line $B$, as $\beta = 1/\alpha$. We see from the definition of meta-lines that the $\theta$-edge crossings of $A$ are the $\phi$-edge crossings of $B$, and similarly the $\theta^{\perp}$-edge crossings of $A$ are the $\phi^{\perp}$ edge crossings of $B$. We only get *both* of these facts because $A$ and $B$ are meta-lines of each other, as the meta-line process takes edge crossings at angles that are close to an integer multiple of $\pi$ (i.e. those that would loop around a cycle before hitting another crossing, or high slopes relative to the cycles), and turns them into edge crossings at angles near $\pi/2$ (so they do not loop through a cycle before hitting another edge crossing, i.e. low slopes). Thus the meta-line process acts on the high slope $\theta^{\perp}$ angles of $A$ and turns them into low slope $\phi$ angles of $B$, and similarly takes high slope $\phi^{\perp}$ angles of $B$ and turns them in to low slope $\theta$ angles of $A$.

We now ask what one line tells us about the other. As mentioned above, crossing a face in a meta-line is the equivalent of crossing an entire cycle, possibly multiple times, in the original line. Since for these two lines, crossing a cycle involves crossing 4 or 5 faces (because $4 < \beta = 2 + \sqrt{5} < 5$), this means that crossing a cycle in the meta-line is the equivalent of crossing 4 or 5 cycles in the original line. Moreover, we can track what face crossing in the meta-line corresponds to which cycle in the original line, and extrapolating from that, we can determine which cycle in the meta-line will correspond to which sequence of cycles in the original line.

For example, $A$ crosses the front face from a $\theta$-edge crossing on the left edge to a $\theta$-edge crossing on the right edge. Then $B$ must have $\phi$-edge crossings at both these points as well. To get from the left edge to the right edge of the front face at angle $\phi$, the line $B$ must go through the cycle front-top-back-bottom. We note that we know it is this cycle, and not front-bottom-back-top, as, given the slope of $A$, the point it hits on the right edge must be $-\sqrt{5} - 2$ above the point on the left edge. As we know $B$ must hit the exact same point on the right edge, it

must travel in the upwards direction to arrive $2+\sqrt{5} \mod 4$, equivalent to $\sqrt{5}-2$ above of where it started. If it went in the downwards direction it would arrive $-(\sqrt{5}-2) \mod 4$ above (i.e. below), which is not equivalent to $-2+\sqrt{5}$, and would thus hit the wrong spot. Using this we can equate each face crossing in the meta-line with a specific cycle in the original line. Extrapolating this as sequence, we can furthermore equate each cycle in the meta-line with a sequence of 4 or 5 cycles in the original line.

It now becomes useful to number the cycles of the cube.

1: front-top-back-bottom
2: front-bottom-back-top
3: front-left-back-right
4: front-right-back-left
5: left-top-right-bottom
6: right-top-left-bottom

(3.5)

Since each cycle contains a combination of specific ways in which a line crosses a face, we can note which cycle in the meta-line will equate to which cycles in the original.

| Meta-line cycle | Original line cycles |
|---|---|
| 1: | 4-5-3-6- |
| 2: | 3-5-4-6- |
| 3: | 1-5-2-6- |
| 4: | 1-6-2-5- |
| 5: | 1-4-2-3- |
| 6: | 1-3-2-4- |

(3.6)

If we were to list out all the cycles of $A$, it would be a combination of the sequences above,

and each cycle would represent a sequence above in $B$, each cycle of which would represent a sequence in $A$, and so on. The first thing to note about this list is, for each cycle, appearing in the meta-line corresponds to every other cycle besides itself and its opposite appearing in the original line, (opposite here means cycle of the same faces in the opposite direction, i.e. the pairs 1 and 2; 3 and 4; 5 and 6). Additionally, no cycle is ever followed by itself or its opposite cycle. This means if $A$, for example, has a 1 cycle, then the next cycle will be either a 3, 4, 5 or 6 cycle. This tells us that $B$ will have at least the cycles 4, 5, 3, 6 with a possible repeat, represented by the 1, and then will have the cycles 1 and 2, from whichever cycle is next, as 3, 4, 5 and 6 all represent a sequence with a 1 and a 2. This shows us that $B$, and similarly $A$, eventually contains all the cycles. Each cycle hits 4 edges in a particular direction, and thus has 4 directed $\theta$ edges. Therefore, as the line will hit each of the 6 cycles, it will hit all of the $2 \times 6 = 24$ directed $\theta$ edges. (This result, while seeming to be fairly obvious, becomes surprisingly important. See Chapter 11 for more details.) We also note that as no cycle is followed by itself or its opposite, there are $6 \times 4 = 24$ possible transitions between cycles; these transitions are each of the 24 $\theta^{\perp}$-edge crossings.

The listing above has one major hindrance in it: it lists the order of cycles that will occur, but not necessarily at the correct starting location. That is, a 1 meta-line cycle can lead to a 4-5-3-6 sequence of cycles, or a 5-3-6-4 sequence of cycles. Moreover, as each cycle represents 4 or 5 cycles, depending on where the cycle starts, so it could be 4-5-3-6-4, or 5-3-6-4-5 as well, though note the order will always be maintained, indicating if there are 5 cycles, it begins and ends with the same cycle. We note that a 1 meta-line cycle can actually represent any of eight different sequences of cycles.

The resolution to this obstacle is the above idea that each cycle transition is a specific $\theta^{\perp}$-directed edge. That is, for example, if a line goes from cycle 1 to cycle 6, it must do so on the right edge of the front face (going in the upper right direction). Similarly, for all 24 of the possible transitions there is exactly one edge direction on which the transition is possible. That means that if the meta-line has a 1 cycle followed by a 6 cycle, we know the 1 cycle ends by crossing the front face, and the 6 cycle begins by crossing the right face, see Figure 3.5. Thus, the sequence of cycles that in the original line, which were indicated by this 1 in the meta-line, must end with cycle the indicated by the meta-line crossing the front face into right face, which

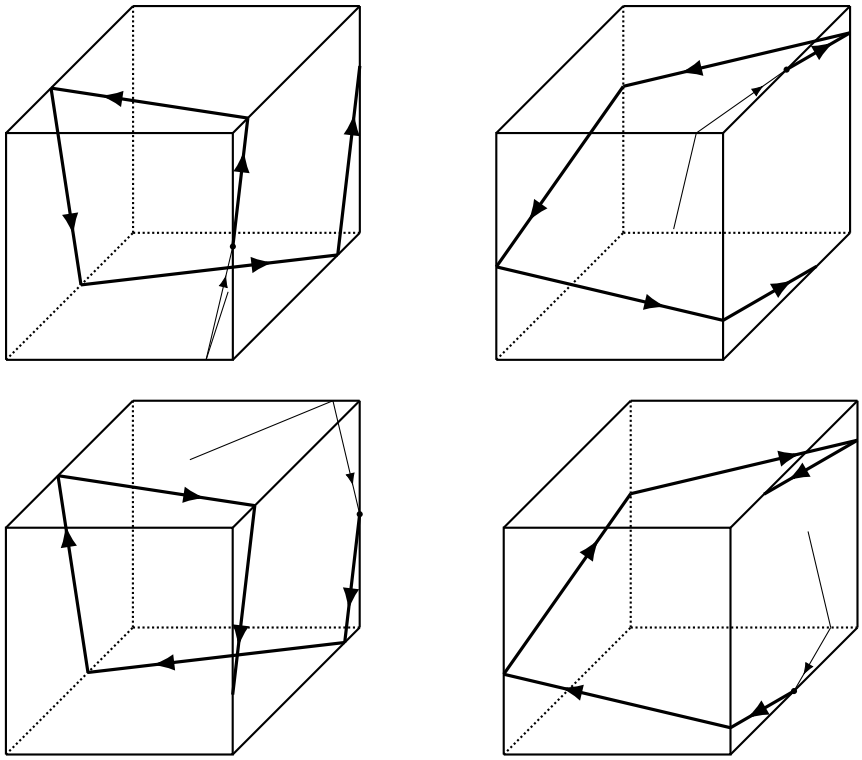

Figure 3.5: Here we have the ending of a 1 cycle (front-top-back-bottom cycle) as it transitions to a new cycle. As we can see which cycle follows is uniquely determined by which edge the 1 cycle ends on.

is the 4 cycle. Similarly, the 6 cycle must begin starting from the right face, and so the sequence of cycles represented by this 6 must begin with the cycle which is represented by the meta-line crossing the right face towards the top face, which is the 1 cycle. Extrapolating this for all possible transitions leads to the chart found in Figure 3.6.

From this it is possible to construct what the lines will look like by providing the sequence of cycles traveled through. We provide an example with $B$.

We can tell from observation that the line $B$ begins with a 1 cycle, and then then a 6 cycle, and that the line $A$ begins with a 4 cycle. From this information we can construct as much of the line as we need using the charts above. $B$ starts off with the cycle sequence

$$B = 16\ldots$$

We know 1 represents 4536, and the sequence must begin with a 4, and 6 represents 1324. A 1-6 transition represents a 4-1 transition, so the first set of cycles must end in a 4, which

| Meta-line transition | Original line transition | Meta-line transition | Original line transition |
|---|---|---|---|
| 1-3 | 6-1 | 2-3 | 5-2 |
| 1-4 | 5-1 | 2-4 | 6-2 |
| 1-5 | 3-1 | 2-5 | 4-2 |
| 1-6 | 4-1 | 2-6 | 3-2 |
| | | | |
| 3-1 | 5-3 | 4-1 | 6-4 |
| 3-2 | 6-3 | 4-2 | 5-4 |
| 3-5 | 2-3 | 4-5 | 1-4 |
| 3-6 | 1-3 | 4-6 | 2-4 |
| | | | |
| 5-1 | 4-5 | 6-1 | 3-6 |
| 5-2 | 3-5 | 6-2 | 4-6 |
| 5-3 | 1-5 | 6-3 | 2-6 |
| 5-4 | 2-5 | 6-4 | 1-6 |

Figure 3.6: Table of which meta-line transition corresponds to which original line transition.

moreover tells use it will be a set of 5 cycles (as it begins and ends with the same cycle), and the second set of cycles must begin with a 1. Thus we know the beginning of $A$ will be

$$A = 453641324\ldots \tag{3.7}$$

We can repeat this process on $A$ to learn more about $B$: 4 represents 1625, and must start with 1, while 5 represents 1423, and a 4-5 transition represents a 1-4 transition. We add on that 3 represents 1526 and a 5-3 transition represents a 1-5 transition. Continuing this we get,

$$B = 162514231526132416251645361526354625 16\ldots \tag{3.8}$$

This process can go on ad infinitum, growing by a factor of $2+\sqrt{5}$ with each step.

What is so interesting about this sequence is that it allows us to actually know what directed edge the line is on at any given time, as each transition in the sequence corresponds to a specific directed edge. As discussed in Chapter 1, much of the difficulty of the problem of the cube line is in not knowing which edge it is on at any given time due to the quasi-random nature of the irrational slopes and non-translatable singular points. With this, however, we can know exactly which directed edge the cube line is on at the $\theta$-edge crossing at $A\left(n\sqrt{1+\beta^2}\right)$, based on the transition of the sequence of cycles for line $B$, calculable in $O(\log(n))$ time, as this sequence grows exponentially.

Moreover, we can relate this sequence to the well known Beatty sequence. The Beatty sequence was first described by S. Beatty in 1926 [1], and denotes the sequence of floor functions of multiples of an irrational number.

Formally, for an irrational number $r$, the Beatty sequence $(B_r(n))_{n=0}^{\infty}$ is defined as

$$B_r(n) = \lfloor rn \rfloor.$$

We can see with this process: it grows at a factor of $\beta = 2+\sqrt{5}$ transitions per iteration so, if the $n^{th}$ iteration has $I_n$ transitions, the number of transitions in the $n+1^{th}$ iteration is

$$I_{n+1} = B_\alpha(I_n)+1,$$

with the $+1$ because it is actually the ceiling function in this case, not the floor as in the classic Beatty sequence.

We further note that the meta-line transitions come in series of length 3. That is, if $B$ has a transition of 1-6, this implies a transition of 4-1 on $A$, which in turn implies 6-4 transition on $B$, which then implies a 1-6 transition on $A$. Repeating this process again tells us that in three more transfers, we will have a $1-6$ on $B$ again. Now we can ask the question: if the first 1-6 is the first transition of the sequence, as it is for $B$ above, when will this 1-6 six implications later occur? It will be the last transition in the sequence after iterating six times, so it will occur at

$$B_\alpha(B_\alpha(B_\alpha(B_\alpha(B_\alpha(B_\alpha(1)+1)+1)+1)+1)+1)+1.$$

Unfortunately, little is known about the sequence of six recursive iterations of the Beatty sequence. That said, it does give, in some way, a solid reference for determining the locations of this quasi-random sequence. Defining

$$B_\alpha^{(i+1)}(0) = (B_\alpha(B_\alpha^{(i)}(0))+1, \quad B_\alpha^{(0)}(0) = 1, \tag{3.9}$$

then we have that all the terms

$$B_\alpha^{(6i)}(0) = 1-6 \text{ transitions } \forall i \in \mathbb{N},$$

which in turn implies $A(\beta * B_\alpha^{(6i)}(0))$ is on the edge direction from the front face to the right face, $\forall i \in \mathbb{N}$.

This gives a general result about an infinite sequence of instances of the cube line hitting a particular edge. Thus we have a simple (albeit poor) quantitative result about the cube line $A$. This technique is further elaborated upon in Chapters 5 and 6, resulting in more substantial quantitative results.

Finally, as a relation to what is known about the Beatty sequence, since we know all transitions will be implied by recursively looking at different portions of earlier iterations of the sequence similar to 3.9, this suggests there might be some extension to Rayleigh's theorem acting here. Note Rayleigh's theorem [19] states that for each irrational number $r$, the number $s = r/(r-1)$ is such that $B_r(n)$ and $B_s(n)$ partition the natural numbers. Similarly, in our case, the different recursions of Beatty's sequence should partition the complete sequence of cycles.

# CHAPTER 4

# EXISTENCE OF DENSE AND UNIFORMLY DISTRIBUTED CUBE LINES

In this chapter we prove that there exists a cube line, which is dense on the cube, and then further is uniformly distributed on the cube, by showing that the cube lines $A$ and $B$ described in Chapter 3 satisfy these requirements. (Recall $A$ is the cube line with initial point a vertex $V$ and initial slope $\sqrt{5}-2$, and $B$ is the cube line with the same initial point and initial slope $\sqrt{5}+2$.) We begin with density.

We first note that, by the arguments in Chapter 2, density of the line on the cube is equivalent to density of the $\theta$-edge crossings of the cube. In fact, the argument for reducing the problem of density on the faces to density of $\theta$-edge crossings on the edges is even simpler than the previous arguments reducing uniformity. It can be seen clearly by noting that for each point on the cube, $a$, and a given line $L(t)$ with angle $\theta$, there is a point on an edge, $b$, that leads to $a$ with slope $\theta$. If a line hits points arbitrarily close to $b$, then it will similarly hit arbitrarily close to $a$. Thus it is enough to show that $T = T_\alpha$ is dense over the set of edges, $[0,24)$.

This process can be simplified a step further to say that it is enough to show density in a single directed $\theta$ edge. This follows from the logic above if we assume that for any point $a$ on the cube, a line going from $a$ in the reverse direction will eventually hit all edges of the cube

(in finitely bounded time). Where it hit on the edge will be used as our $b$ point and the logic above follows. As mentioned in Chapter 3, lines $A$ and $B$ do indeed hit all directed edges, and in fact, as we will see later, all cube lines with irrational slopes will hit all directed edges (see Lemma 16 and Lemma 18).

We begin by looking at the line $A$ and considering where on each directed $\theta$ edge each $\theta$-edge crossing occurs. Let $a_i = A\left(i\sqrt{1+\alpha^2}\right) \in [0,1)$ be the location where the $n^{th}$ $\theta$-edge crossing occurs. If one $\theta$-edge crossing occurs at $a_i \in [0,1)$ on a directed edge, then the next will occur at $a_i + 1/\beta = a_0 + \alpha$ modulo 1, on some other directed edge, and thus the sequence of points on the directed edges where the $\theta$-edge crossings occur is the irrational rotation,

$$a_0 + n\alpha \mod 1 \quad n \in \mathbb{N}.$$

As mentioned in Chapter 1, the properties of the irrational rotation are well understood, namely that it is dense and uniformly distributed. Thus the sequence of locations of $\theta$-edge crossings with their directed $\theta$ edges, $(a_n)_{n=0}^{\infty}$, is a dense, uniformly distributed set on the half open interval $[0,1)$, because irrational rotations are dense and uniformly distributed. Therefore, we can think of the set of $\theta$-edge crossings as taking this dense set (derived from the sequence) of the half open interval $[0,1)$, and partitioning its elements among the 24 directed $\theta$ edges by putting each $\theta$-edge crossing in the partition piece of whatever directed edge it occurs in.

From here we show a general lemma about dense subsets of intervals.

**Lemma 3.** *Let $I$ be a real interval and $D$ a dense set in $I$. For any finite partition $\Gamma$ of $D$, there is a non-empty open sub-interval $U \subseteq I$ and element $\gamma \in \Gamma$ such that $\gamma$ is dense in $U$.*

Proof by induction on $n = |\Gamma|$, the number of partition pieces in the partition $\Gamma$. The base case of $n = 1$ is trivial as $\gamma \in \Gamma = D$ is dense in $I \subseteq I$. Let $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$. Suppose $\gamma_n$ is not dense in $I$. Then by definition there is a non-empty open interval $I_0 \in I$ such that $\gamma_n \cap I_0 = \varnothing$. Therefore $D \cap I_0$ has been partitioned into $n-1$ elements by $\Gamma \cap I_0$. Applying the inductive hypothesis implies there is some open interval in $I_0$, and thus in $I$, for which the condition holds. □

Applying this lemma to the set of $\theta$-edge crossing locations within their directed edges,

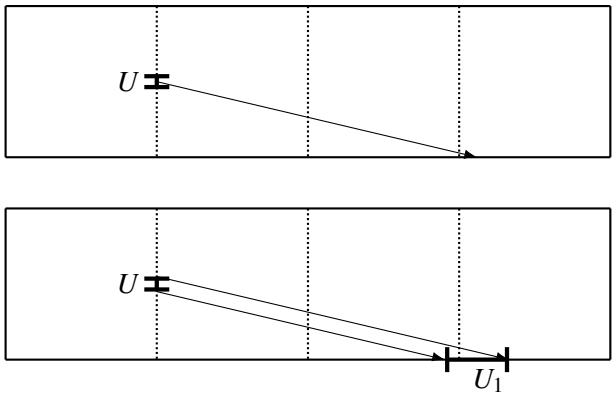


Figure 4.1: Using the initial dense set $U$, we can extrapolate the existence of a larger dense set $U_1$. Note we only take the portion of the new set entirely contained in one edge (opting for the larger selection).

which are dense in the interval $[0,1)$, partitioned by which directed $\theta$ edge they occur in, we get that there is some non-empty open interval $(r_1,r_2) \subset [0,1)$ such that one piece of the partition is dense on $(r_1,r_2)$. This means there exists some sub-interval of some directed $\theta$ edge in which the $\theta$-edge crossings are dense.

Let $U$ be an open interval on a directed edge $\vec{\mathscr{E}}$ for which the points

$$\left\{A\left(n\sqrt{1+\alpha^2}\right)\right\}_{n=0}^{\infty}$$

are dense; we refer to the set of $\theta$-edge crossings in $U$ as $U_0$. Since $A$ and $B$ hit the edges at the same points, $B$ will also hit all points in $U$, and thus the dense set $U_0$ is also in $B$, where they are $\phi$-edge crossings (again, $\phi$ is the angle of $B$). The idea of the following procedure to use the fact that this set is dense to show that a slightly bigger interval in which $A$ is dense, which in turn will show a slight bigger interval, and so on, until we show that an entire directed edge must have a dense set in $A$.

Consider a point $x_0 \in U_0$, a $\theta$-edge crossing of $A$, and let $(v_n)_{n=0}^{\infty}$ be the in order sequence $\theta^{\perp}$-edge crossings of $A$. Then $\exists i \in \mathbb{N}$ such that $x_0$ occurs between $v_i$ and $v_{i+1}$, that is, $A(t_1) = v_i$, $A(t_2) = x_0$ and $A(t_3) = v_{i+1}$, with $t_1 < t_2 < t_3$, and we select the point $v_{i+1}$, that is, we take the $\theta^{\perp}$-edge crossing which occurs most immediately after $x_0$. We call $v_{i+1}$ the *image* of $x_0$. We do this for all $x \in U_0$, and call the set of image $\theta^{\perp}$-edge crossings $U_1$.

We claim $U_1$ is dense in an interval. Consider the cycle between $v_i$ and $v_{i+1}$. Note that this

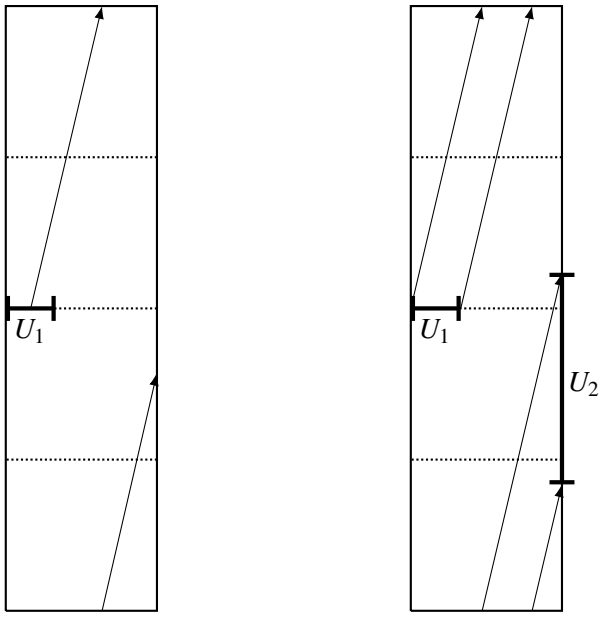


Figure 4.2: Iterating this process with the new set $U_1$ yields increasingly large dense sets, until eventually we have a dense set which contains an entire directed edge.

cycle must be the same cycle for all $x \in U_0$, as all the points in $U_0$ are $\theta$-edge crossings on the same directed edge. For simplicity, assume the cycle is the 4 cycle (the front-right-back-left cycle) and $U$ is on the edge between the front and right faces. Then the $\theta^{\perp}$-edge crossing $v_i$ is the wherever the line hits on the top edge of one of the faces, immediately after hitting $x_0$. It is quite clear from looking at the cycle as a 1 by 4 torus that this mapping from $x_0$ to $v_{i+1}$ is a continuous, linear mapping, and will in fact be the same linear mapping for all points in $U_0$. The continuous mapping of a dense set on an interval will be a dense set on an interval, thus $U_1$ will be dense on an interval. Note $U_1$ might be split over more than one directed edge. In this case, we take the largest interval on any one edge and call it $U_1$, see Figure 4.1.

We next look at the size of $U_1$ (size here refers to interval length). Take two points of $U$, $x_1$ and $x_2 \in U \subset [0,1)$ (denoted by their location on the directed $\theta$ edge ), and suppose $|x_1 - x_2| = \varepsilon > 0$. We label their images $y_1$ and $y_2$ respectively (denoting the image by their location in the directed edge containing $U_1$). Then we have, noting $-2+\sqrt{5} = \frac{1}{2+\sqrt{5}}$ is the slope of $A$,

$$y_1 \bmod 1 = (1-x_1)(2+\sqrt{5}), \quad y_2 \bmod 1 \ = (1-x_2)(2+\sqrt{5}).$$

Without loss of generality, we can assume $x_2 = x_1 + \varepsilon$, so we have

$$y_2 \bmod 1 \ = (1 - x_1 - \varepsilon)(2 + \sqrt{5})$$

$$y_2 \bmod 1 \ = y_1 \bmod 1 \ - \varepsilon(2 + \sqrt{5})$$

Assuming $x_1$ and $x_2 \in U_0$ are close enough together, their images will be on the same directed edge and thus we can drop the modulo 1 on both of them. This gives us,

$$y_2 - y_1 = (2 + \sqrt{5})\varepsilon.$$

Thus, for two points in $U$, their images in $U_1$ are $2 + \sqrt{5}$ times further apart. Thus we now have a much larger dense set, $U_1$. If $U_1$ contains an entire edge, we have an density in an entire directed edge, and thus we are done. If $U_1$ does not contain an entire edge, then we have $|U_1| = (2 + \sqrt{5})|U|$ (again, here $|.|$ refers to interval length), noting that not containing an entire edge prevents any overlapping in the linear map that defines $U_1$. It is possible $U_1$ spans more than 1 edge, so we pick the largest part entirely on one edge, which we know is at least of size $2|U|$ (this follows from the fact that $U_1$ does not contain a whole edge, and thus the interval it is dense in must be contained in at most two directed edges).

We can now repeat this process, using $B$. As $U_1$ is $\theta^{\perp}$-edge crossings in $A$, it is $\phi$ edge crossings in $B$. As $\theta = \phi^{\perp}$, doing this process again will give us a set of image $\theta$-edge crossings, $U_2$, that either contains a full edge, or a dense set on an edge of size at least $4|U|$, see Figure 4.2. We continue this process, alternating between finding $\theta^{\perp}$-edge crossings from $\theta$-edge crossings, and $\phi$-edge crossings from $\phi^{\perp}$-edge crossings. Repeating this process at most $\log_2 \frac{2}{|U|}$ times yields a set of images which must contain a complete edge. This is a set which is dense on an entire edge-direction, thus proving the following theorem.

**Theorem 3.** *Let A and B be two cube lines such that A is the meta-line of B and B is the meta-line of A (see 3.1 for explicit formula). Then both A and B are dense on the cube.*

□

We next aim to prove that the cube lines $A$ and $B$, and all those described in 3.1, are uniformly distributed. For simplicity we work on line $A$.

We begin by constructing a slightly different transformation than that of $T_\alpha$ described in Chapter 2. Consider the set of half open unit intervals $[0,1)$ representing each of the directed $\theta$ edges and each of the directed $\theta^\perp$ edges, and let $\Omega = [0,48)$ be the interval of them glued together in the obvious way (see 2.2). We chose to represent this as $\Omega = E \times M$, where $E = [0,1)^{24}$ represents the directed $\theta$ edges of $A$, and $M = [0,1)^{24}$ represents the directed $\theta^\perp$ edges for the line $A$.

We also define the mapping $\Psi : \Omega \to [0,1)$ in the obvious way, taking each half open interval $[r, r+1) \in \Omega$, for integers $0 \leq r < 48$, shrinking it linearly by a factor of 48, and then placing them in sequence on the unit interval. Formally we can define the function $\Psi$ as follows: let $\Omega = \{[0,1)_i : 0 \leq i < 48\}$ be set of edges with some arbitrary sequencing. Then we define $\Psi$ by

$$\Psi : \{[0,1)_i : 0 \leq i < 48\} \to [0,1)$$

$$\Psi(x_i) = \frac{i}{48} + x_i, \quad \forall x_i \in [0,1)_i,\ \forall 0 \leq i < 48. \tag{4.1}$$

Next we define two mappings on this space, $\Omega$. First, we define a mapping on the half-space, $E$; let the mapping $\tau = \tau_A : E \to E$ be defined as follows: for any $x \in E$, assume a geodesic $A$ with angle $\theta$ is at the point $x$ traveling in the direction of the directed $\theta$ edge that $x$ is located on, and let $\tau x$ be the next point in $E$ which the line hits. We note that by construction, $\tau_A = T_\alpha$, where $T_\alpha$ is our standard mapping as defined in Chapter 2. Thus, by the results in Chapter 2, if the orbit of a point under function $\tau$ is uniformly distributed over the space $E$, then the line $A$ will be uniformly distributed on the edge directions of the cube, and thus uniformly distributed on the cube.

We next define a second mapping $G = G_A : \Omega \to \Omega$. For any $x \in E$, we again assume a geodesic $A$ with angle $\theta$ is at the point $x$ traveling in the direction of the directed $\theta$ edge thaat $x$ is located on, and define $Gx$ to be the next point in $M$ which the line hits—that is, the next $\theta^\perp$-edge crossing the geodesic $A$ would hit. For $x \in M$ however, we make use of the fact that the directed $\theta^\perp$ edges of a line $A$ are the directed $\phi^\perp$ edges of its meta-line, $B$. (Again, recall $\theta^\perp = \phi$). So for $x \in M$ we instead assume a geodesic $B$ with angle $\phi$, is at the point $x$ traveling in the direction of the directed $\phi^\perp$ edge of $B$ that $x$ is located on, and define $Gx$ to be the next

point in $E$ which the line hits. That is, we send each $\phi$-edge crossing to its next $\phi^{\perp}$-edge crossing. Note the definition of $G$ is essentially the same as the back and forth process used early in the chapter to Theorem 3, see Figures 4.1 and 4.2.

The basic idea of our proof is to show that the transformation $G$ is ergodic (though it is **not** measure preserving), then prove that this fact implies $\tau$ is ergodic (which is measure preserving), and finally that $\tau$ is ergodic proves that $\tau$ is uniformly distributed for almost every starting location.

We remind the reader that a transformation $G$ is called *ergodic* if the only sets which it holds constant are have either empty measure or full measure. That is $G$ is ergodic if, for any set $A$ satisfying $GA = A$, either $A$ has measure zero or $A^c$, the compliment of $A$, has measure zero. Note this definition is independent of whether $G$ is measure preserving or not.

**Lemma 4.** *The mapping $G$ is ergodic over $\Omega$ (with respect to the 1-dimensional Lebesgue measure).*

Using results by published by K. Wilkinson [21] in 1975, for a transformation $T$ which is a piecewise linear function from $[0,1)$ to itself, we can use the following conditions to imply the function is ergodic:
Let

$$b_i > 0, \qquad \sum_{i \in I} b_i = 1$$

$$a_0 = 0, \qquad a_{i+1} = a_i + b_i$$

$$Tx = c_i(x - a_i) + d_i,$$

if $c = \inf_{i \in I} c_i > 1$ and for each $i \in I$, $c_i b_i + d_i \leq 1$, with $d_i > 0$, then $T$ is ergodic.

We claim that $h = \Psi \circ G \circ \Psi^{-1}$ satisfies these conditions. First we show that $G$ is piecewise linear and thus $h$ can be written as above. Let $Y = (a,b) \subseteq [0,1)$ be an open set in one of the directed edges of $\Omega$. If $Y \subset E$ then $GY$ is the set $\theta^{\perp}$-edge crossings which are next hit by a line of angle $\theta$ starting from the points in $Y$. Without loss of generality suppose $Y \subset E$ is on

the directed $\theta$ edge going from the left face to the front face. Again this can be achieved by rotating or mirroring the cube if needed. If we look at this edge as a whole, we note that the interval
$[1-\alpha, 1)$ maps linearly to the front top edge. This can be seen clearly by looking at the front-right-back-left cycle as a 1 by 4 torus line. Similarly the interval
$[1-2\alpha, 1-\alpha)$ will map linearly to the right top edge, $[1-3\alpha, 1-2\alpha)$ to the back top edge, $[1-4\alpha, 3\alpha)$ to the left top, and finally $[0, 1-4\alpha)$ will map linearly to
$[1-\alpha(1-4\alpha), 1)$ (note these are not measure preserving linear mappings). Thus we can see that the interval $M$ will embed into these 5 linear maps and be a piecewise linear map with at most 4 points of discontinuity, $1-\alpha$, $1-2\alpha$, $1-3\alpha$, and $1-4\alpha$. Generalizing this same argument to all directed $\theta$ edges and directed $\theta^{\perp}$ edges we see that $G$ is piecewise linear. As $\Psi$ is also a piecewise linear mapping, mapping each edge linearly to an interval $[n/48, (n+1)/48)$ for some $0 \leq n < 48$, this implies $h$ is also a piecewise linear map and can thus be written as above.

We next note that each portion of the edge is linearly mapped to an interval that is $1/\alpha = \beta$ times larger. Thus we note that all the linear coefficients $c_i = \beta$ for all $i$, as it is clear conjugation by $\Psi$ will not affect the linear scaling.

Lastly we need to show $c_i b_i + d_i \leq 1$. But as

$$a_{i+1} - a_i = b_i$$

and

$$Tx = c_i(x - a_i) + d + i$$

for $x \in [a_i, a_{i+1})$, this is simply saying that all linear pieces have their images fitting inside the entire space, i.e. the function has been taken modulo 1. It is clear that all linear pieces of $G$ fit within the entire space, and conjugation by $\Psi$ ensures that the function is equivalent modulo 1. Thus $h$ satisfies all of the above conditions and thus $h$ is ergodic.

To complete the lemma we must finally confirm that $h$ being ergodic implies $G$ is ergodic. But we note that $\Psi$ is a piecewise linear bijection. Suppose $Y \subset \Omega$ such that $GY = Y$. Then $h(\Psi(Y)) = \Psi(Y)$. Thus $\mu(\Psi(Y)) = 0$ or 1. As $\Psi$ is a piecewise linear bijection with a linear

factor of $1/48$, this implies $\mu(Y) = 0$ or 48, thus $\mu(Y) = \mu(\varnothing)$ or $\mu(Y) = \mu(\Omega)$, thus $G$ is ergodic over $\Omega$. □

**Lemma 5.** *The mapping $\tau$ is ergodic over $E$ (with respect to the 1-dimensional Lebesgue measure).*

This lemma is a result of the above note: a line $A$ hits a $\theta$-edge crossing at a point $x$, if and only if its meta-line hits a $\phi$-edge crossing at the same point $x$.

Let $F \subset E$ be such that $\tau F = F$. Let $\mathscr{F} \subset \Omega$ be the smallest subset of $\Omega$ such that $F \subset \mathscr{F}$ and $G\mathscr{F} = \mathscr{F}$ (note we know this minimum exists as it is equal to the intersection of all such subsets, and the fact $\mathscr{F} = \Omega$ satisfies shows the intersection is non-empty). As $G$ is ergodic this implies $\mu(\mathscr{F}) = 0$ or $\mu(\mathscr{F}) = \mu(\Omega) = 48$. If $\mu(\mathscr{F}) = 0$ this implies $\mu(F) = 0$ and we are done. If $\mu(\mathscr{F}) = \mu(S) = 48$, this implies that $\mu(\mathscr{F} \cap E) = 24$.

As $\mathscr{F}$ is the smallest of the sets contain $F$ which are their own image, we have that

$$\mathscr{F} = \bigcup_{n=0}^{\infty} G^n F. \tag{4.2}$$

Thus for all $y \in \mathscr{F} \cap E$, $y = G^n x$ for some $n \in \mathbb{N}$ and some $x \in F$. As $x \in E$, $x$ is a $\theta$-edge crossing hit by a line $A$ with angle $\theta$ starting at $x$, and also a $\phi$-edge crossing of its meta-line $B$ with angle $\phi$ starting at $x \in E$. It then hits a point in $M$ at $Gx \in M$, a $\theta^{\perp}$-edge crossing of $A$ but also a $\phi^{\perp}$-edge crossing of $B$. Then it follows, $G^2(x) \in E$ is some $\phi$-edge crossing of $B$, and is therefore also a $\theta$-edge crossing of $A$. This logic follows until we get to $y = G^n x \in E$ must be some $\theta$-edge crossing of $A$. But the set of all $\theta$-edge crossings of the line $A$ is the set $\{\tau^n x : n \in \mathbb{N}\}$. As the set $F$ is its own image, this implies

$$\bigcup_{n=0}^{\infty} \tau^n F \subseteq F \Rightarrow \tau^n x \in F \Rightarrow y \in F.$$

Thus we have

$$\mathscr{F} \cap E \subseteq F \Rightarrow \mu(F) \geq \mu(\mathscr{F} \cap E)$$

$$\Rightarrow \mu(F) \geq 24 \Rightarrow \mu(F) = 24 = \mu(E).$$

Therefore, if $\tau F = F$, then $\mu(F) = 0$ or $\mu(F) = \mu(E)$, and thus $\tau$ is ergodic over $E$, proving the lemma. □

The final step is to show that $\tau$ being ergodic over $E$ implies that for almost every $x \in E$, $\{\tau^n(x)\}_{n=0}^{\infty}$ is uniformly distributed over $E$.

This is a direct result from Birkhoff's ergodic theorem [3].

We give a brief review of Birkhoff's Theorem for measure preserving flows[3]: Let $(\chi, \mathscr{A}, \mu)$ be a probability space; and $T_t : X \to X$, $t \in \mathbb{R}$ be an invertible measurable flow, meaning a measure-preserving group action $T_{s+t} = T_t T_s$; and let $f \in L^1(\chi, \mathscr{A}, \mu)$ be an integrable test function. Then we have

$$\lim_{N\to\infty} \frac{1}{N} \int_0^N f(T_t x)dt = \lim_{N\to\infty} \frac{1}{N} \int_0^N f(T_{-t} x)dt = \bar{f}(x) \tag{4.3}$$

for almost every $x \in \chi$, with respect to measure $\mu$, and:

1. $\bar{f}(T_t x) = \bar{f}(x)$ almost everywhere—that is, the limit function $\bar{f}$ is $\{T_t\}$-invariant;

2. the flow is called ergodic if every $\{T_t\}$-invariant measurable set $A \in \mathscr{A}$ with $T_t(A) = A$ almost everywhere has the property $\mu(A) = 0$ or $\mu(A) = 1$;

3. if the flow $(T_t)$ is ergodic then every $(T_t)$-invariant function is constant, and the right-hand side of 4.3 simplifies to the integral

$$\int_{\chi} f(x)d\mu(x).$$

A well known example of an ergodic flow is the irrational torus line flow,

$$T_t : x \to x + t\mathbf{v} \text{ modulo } 1,$$

where $\mathbf{v} = (v_1, v_2)$ is fixed with an irrational slope $v_1/v_2$, and $x$ runs over the 2-dimensional torus $[0,1)^2$.

As $\tau$ is a measure preserving mapping, 4.3 implies the orbit of almost any point $x$ under it is equidistributed. A proof of this standard result is as follows.

Birkhoff's ergodic theorem tells that if $\tau$ is an ergodic transformation in $E$, then for any measurable function $f$, for almost every $x \in E$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n} f \circ \tau^j(x) = \int_E f dm.$$

Let $\mathbb{Q}_1$ be the set of rational numbers in the interval $[0,1)$. Let $\mathbb{E}$ be the set defined as

$$\mathbb{E} = \{(x,y;z) : x,y \in \mathbb{Q}_1, x < y, z \in \mathbb{N}, 0 \leq z < 24\}.$$

This set represents all open intervals with rational endpoints in $E$, with the $z$ representing which directed edge it is located on. We note that this is countable so we let $(\gamma_i)_{i=0}^{\infty}$ be some enumeration of the set $\mathbb{E}$. We then define $f_{\gamma_i}$ to be the indicator function for the interval represented by $\gamma_i$.

For each $f_{\gamma_i}$, we know that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n} f_{\gamma_i} \circ \tau^j(x) = \int_E f_{\gamma_i} dm$$

holds for almost every $x$. Call the set it does not hold for $P_{\gamma_i}$. Note $\mu(P_{\gamma_i}) = 0$ for all $\gamma_i$, and as there are a countable number of $\gamma_i$, this implies,

$$\mu\left(\bigcup_{i=0}^{\infty} P_{\gamma_i}\right) \leq \sum_{i=0}^{\infty} \mu\left(P_{\gamma_i}\right) = \sum_{i=0}^{\infty} 0 = 0.$$

Thus for almost every $x \in E$, the orbit of $\tau$ is uniformly distributed for every interval with rational endpoints. Let $x$ be any point not in $\bigcup_{i=0}^{\infty} P_{\gamma_i}$. Let $(a,b)$ be any interval on a directed $\theta$ edge of $E$. For any $\varepsilon > 0$, there exists rational numbers $a_\varepsilon^+$, $b_\varepsilon^-$, $a_\varepsilon^-$, and $b_\varepsilon^+$ in the interval $[0,1)$, such that,

$$0 \leq a_\varepsilon^+ - a \leq \varepsilon,$$

$$0 \leq b - b_\varepsilon^- \leq \varepsilon,$$

$$0 \leq a - a_\varepsilon^- \leq \varepsilon,$$

$$0 \leq b_\varepsilon^+ - b \leq \varepsilon.$$

That is, $a_\varepsilon^+$ and $b_\varepsilon^+$ are slightly greater than $a$ and $b$, respectively, and $a_\varepsilon^-$ and $b_\varepsilon^-$ are slightly less than $a$ and $b$, respectively.

Further, letting $f_I$ be the indicator function for any interval $I$, we know that

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a,b)} \circ \tau^j(x) \geq \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^+, b_\varepsilon^-)} \circ \tau^j(x)$$

$$= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^+, b_\varepsilon^-)} \circ \tau^j(x) = \int_E f_{(a_\varepsilon^+, b_\varepsilon^-)} dm = b_\varepsilon^- - a_\varepsilon^+ \geq b - a - 2\varepsilon$$

and

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a,b)} \circ \tau^j(x) \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^-, b_\varepsilon^+)} \circ \tau^j(x)$$

$$= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^-, b_\varepsilon^+)} \circ \tau^j(x) = \int_E f_{(a_\varepsilon^-, b_\varepsilon^+)} dm = b_\varepsilon^- - a_\varepsilon^+ \leq b - a + 2\varepsilon.$$

Putting these together we have that

$$b - a - 2\varepsilon \leq \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^+, b_\varepsilon^-)} \circ \tau^j(x) \leq$$

$$\leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a,b)} \circ \tau^j(x) \leq b - a + 2\varepsilon.$$

Thus, as $\varepsilon$ can be made to be arbitrarily small, taking the limit as $\varepsilon$ goes to 0 we get

$$\lim_{\varepsilon\to 0} b - a - 2\varepsilon \leq \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^+, b_\varepsilon^-)} \circ \tau^j(x)$$

$$\leq \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a,b)} \circ \tau^j(x) \leq \lim_{\varepsilon\to 0} b - a + 2\varepsilon$$

$$\Rightarrow b - a \leq \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^+, b_\varepsilon^-)} \circ \tau^j(x) \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a,b)} \circ \tau^j(x) \leq b - a$$

$$\Rightarrow b - a = \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a_\varepsilon^+, b_\varepsilon^-)} \circ \tau^j(x) = \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} f_{(a,b)} \circ \tau^j(x)$$

$$\Rightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n} f_{(a_\varepsilon^+, b_\varepsilon^-)} \circ \tau^j(x) = b - a.$$

Thus proving that the orbit of $x$ under $\tau$ is uniformly distributed over $E$, and thus, for almost every $x \in E$, the orbit under $\tau$ is uniformly distributed over $E$. Therefore, using Lemma 1 and Lemma 2 we have proved the following theorem about the existence of a uniformly distributed cube line on the surface of a cube.

**Theorem 4.** *Let A be any cube line with slope given by 3.3.*

1. *For almost every starting point $a_0$, A is uniformly distributed over the surface of the cube in the Weyl sense (see Theorem 1).*

2. *For almost every starting point $a_0$, A is uniformly distributed over the surface of the cube in the Birkhoff sense (see Theorem 2).*

□

# CHAPTER 5

# QUANTITATIVE RESULTS: COMBINATORIAL UNIFORMITY

The proofs of Theorems 1, 2, 4,9-12 are all based on the same two-step approach: (1) show that the corresponding measure-preserving transformation is ergodic, (2) use Birkhoff's well known ergodic theorem. The main detriment to using Birkhoff's ergodic theorem is that it provides no insight to any quantitative results with regard to uniform convergence—that is, it tells us nothing about the rate at which uniform convergence occurs, just that it does indeed occur as the limit goes to infinity. Previous work on uniformity of torus lines on a square and billiard paths on a square proved uniform distribution using Weyl's criterion, which has several quantitative versions (e.g. the Erdős-Turán-Koksma inequality [5]); Birkhoff's ergodic theorem, however, has no such quantitative version. Thus to find any quantitative results about uniformity we must look elsewhere. In this chapter we formulate some quantitative results for the geodesic in the special case of cube lines $A$ and $B$, as describe in the previous chapters.

We first look at uniformity in a slightly different way; we label the 6 faces of the cubes with numbers 1,2,3,4,5,6, and write down the sequence of faces visited by a particle moving on cube line $L(t)$ $t \to \infty$: we call this the "face-crossing" sequence. This infinite sequence will be aperiodic (assuming irrational slope) and will exhibit "pseudo-random" behavior. In comparison to textbook "randomness", there is a clear difference between the face-crossing sequence and true randomness in their local aspects. For example, a fair die with the numbers 1 through 6

being rolled to generate an infinite sequence will have long runs (say) ... 1111111111 ..., but the face-crossing sequence will never even have a run of length two, i.e. every face-crossing is followed by a face-crossing of a different face.

Despite this dissimilarity in the local aspects, the global aspects of the two sequences are much more similar. The law of large numbers implies that in a typical random die sequence, every integer $k$, $1 \leq k \leq 6$ has the same asymptotic density 1/6, and our results from Theorem 4 imply the same is true for the face-crossing sequence. We denote this property by saying the face-crossing sequence exhibits *combinatorial uniformity*. The same holds for the "edge-crossing" sequences, and "directed edge-crossing" sequence, with the straight forward change that the asymptotic density is 1/12 and 1/24, respectively.

One important quantitative global parameter is the size of the maximum fluctuation around the mean value. It is well know that in a typical random die sequence the maximum fluctuation exhibited is roughly square-root sized—in fact, for a random die sequence the typical fluctuation and maximum fluctuation turn out to be nearly the same [11]. This leads us to ask a similar question regarding the face-crossing sequence: what is the size of the maximum fluctuation of the face-crossing sequence?

We start by looking at some previously known results of a simpler analogous question: What is the size of the analog maximum fluctuation in the square billiard orbit? As the square billiard orbit has only one face, the analogous question is rather about edge-crossing. To be precise, we consider two parallel edges, say the top and bottom edges, of the square. Consider $N$ consecutive horizontal edge-crosses of the square billiard orbit (that is, whenever the billiard orbit "bounces" off the top or bottom edge). Let $Top(N)$ and $Bottom(N)$ denote the number of crosses of the top and bottom edges respectively. We know trivially that

$$Top(N) = \frac{1}{2}N + o(N) \text{ and } Bottom(N) = \frac{1}{2}N + o(N).$$

But what can we say about the asymptotic behavior of the difference,

$$|Top(N) - Bottom(N)|,$$

as $N \to \infty$? According to the classic works of Hardy, Littlewood, Ostrowski, König and Szücs, the answer depends on the continued fraction digits of the slope [9][17][14]. The continued fraction representation is completely characterized for a single class of irrational numbers—namely the quadratic irrationals are exactly those with periodic continued fractions. It turns out that for every quadratic slope the maximum fluctuation of $|Top(N) - Bottom(N)|$ is in the range of $\log N$. Logarithmic fluctuation is much smaller than the square root fluctuation found in the random die sequence. Moreover the typical fluctuation of $|Top(N) - Bottom(N)|$ is in the range of $\sqrt{\log N}$, much smaller than the maximum fluctuation, again demonstrating a sizable difference between the edge-crossing sequence's fluctuations and that of the random die sequence.

Returning to our question of the face-crossing sequence, we can see there is no immediately clear answer to maximum fluctuation size: is it still negligible logarithmic like the parallel edge-crossing sequence of the square billiard orbit, or roughly $N^{1/2}$ like the random die sequence, or roughly $N^c$, $0 \leq c \leq 1,\ c \neq 1/2$? Our goal is to answer this quantitative question regarding combinatorial uniformity for some particular cube lines.

First we show that the quantitative aspects of combinatorial uniformity for cube line $A$, i.e. cube line starting at a vertex with slope $5 - \sqrt{2}$, depends (quite surprisingly) on the second largest eigenvalue of the corresponding "transition matrix". Then we show how quantitative combinatorial uniformity of $A$ implies quantitative measure-theoretic uniformity of the cube lines of the same slope and arbitrary starting point.

We return to the process of identifying the sequence of cycles of $A$ (and $B$) in Chapter 3. Let us jump into the middle of obtaining 3.8 from 3.7. Consider the concrete consecutive triple 641 in the middle of $A = 453(641)324\ldots$. Again applying 3.6 and Figure 3.6, we know 6-4 $\Rightarrow$ 1-6 and $4 \longrightarrow 1-6-2-5$, and thus the cycle corresponding to 4 must begin with 5, so it is either 6251 or 62516. Since 4-1 $\Rightarrow$ 6-4 and $4 \longrightarrow 4-5-3-6$, the cycle corresponding to 4 has to be 62516, and the cycle corresponding to 1 must begin with 4, so it is either 4536 or 45364. Therefore, simply knowing the partial information $A = \ldots 641 \ldots$ anywhere in the middle of $A$, we can guarantee the corresponding part $B = \ldots 625164526 \ldots$ in the middle of $B$. We say that

$$516,164,645,452,526$$

are the new consecutive triples generated (in $B$) by the triple 641 in the middle (of $A$).

To be clear as to why 625 and 251 are missing from this list of triples despite being in the sequence 625164526; the triples 625 and 251 are not new—they were already generated by the triple 364, the left neighbor of the triple 641 in
$A = 45[3(64]1)324\ldots$. Indeed we can find the list of triples generated by 364 using the same method as above. We have 3-6 $\Rightarrow$ 1-3 and $6 \longrightarrow 1-3-2-4$, so we know that the cycle which corresponds to 6 must begin with 3, and is thus either 3241 or 32413. As again 6-4 $\Rightarrow$ 1-6 and $4 \longrightarrow 1-6-2-5$, the cycle corresponding to 6 has to be 3241, and the cycle corresponding to 4 must begin with 6, so it is either 6251 or 62516. Thus the partial information $A = \ldots 364\ldots$ somewhere in the middle of $A$ guarantees the corresponding part $B = \ldots 32416251\ldots$ in the middle of $B$. The new triples generated by 364 are

$$416, 162, 625, 251.$$

So, 641 generates 5 new triples, and 341 generates 4 new triples.

We can extend this method to all possible (legitimate) consecutive triples, which results in an explicit $k \times k$ "transition matrix" (similar to that of Markov chains), where $k$ is the number of (legitimate) consecutive triples. If we specify the beginnings: say, $A = 452\ldots$ and $B = 162\ldots$, then the rest of the consecutive triples are all in the middle, and the transition matrix applies.

The powers of the explicit transition matrix gives an explicit formula for the number of consecutive triples (of cycles) in terms of the eigenvalues. For example, after $n$ iterations, the number of consecutive triples of cycles 641 (or 364, or any other consecutive triple) is expressed in terms of the $n^{th}$ powers of the eigenvalues. From these explicit formulas it is easy to get a similar explicit formula for the number of crossings of a given edge, or the number of visits of a given face of the cube.

The largest eigenvalue of the $k \times k$ transition matrix is expected to be $2+\sqrt{5}$, so the critical new information is the size of the second largest eigenvalue. From here we have 3 possibilities: (1) the second largest eigenvalue happens to be $\sqrt{2+\sqrt{5}}$, which would imply square-root fluctuation (as in the random die sequence) and thus exhibit "randomness"; (2) the second largest eigenvalue is not $\sqrt{2+\sqrt{5}}$, and it has absolute value greater than one, which would mean

a new kind of "non-random chaos"; (3) the second largest eigenvalue has absolute value not greater than 1, which means the exponent term would be negligible as $N \to \infty$, and thus imply negligible fluctuation as in the case of the square billiard orbit. We will show that the second largest eigenvalue is greater than $\sqrt{2+\sqrt{5}}$, implying case (2).

The primary challenge with showing this is that the transition matrix of consecutive triples is quite large, and thus quite difficult to compute the eigenvalues for. Counting the number of legitimate consecutive triples, there are 6 possibilities for the first cycle, which can end on one of 4 edges resulting in 4 possibilities for the second cycle, which given the edge it began on has only 2 possibilities for which end edge it can end on, resulting in 2 possibilities for the third cycle, or $k = 48$ cycles total.

Due to these large number of triples, we introduce a variable like $x_n(641)$ and $x_n(364)$. Given the cycle these triples generate, we can then write out the system of simultaneous recurrences in the following manner:

$$x_{n+1}(641) = x_n(516) + x_n(164) + x_n(645) + x_n(452) + x_n(526)$$

and

$$x_{n+1}(364) = x_n(416) + x_n(162) + x_n(625) + x_n(251).$$

This still leaves us with a very unwieldy $48 \times 48$ matrix to find the eigenvalues of. We can however simplify this process substantially by reducing the number of variables via symmetry.

We make the following reduction: We identify points of the cube according to the 3 mirror-symmetries of the cube: reflection in the $xy$-plane, reflection in the $xz$-plane and reflection in the $yz$-plane, assuming the cube is centered on the origin. Doing this we lose some information, but not much. It means we are actually studying measure-theoretic uniformity relative to a class of partially symmetric test sets. For example, taking a small set test set near a vertex and on a single face of the cube, the 3 mirror symmetries generate altogether $2^3 = 8$ congruent copies of $S$ (4 copies on the original face, and 4 copies on the opposite face). The union of these 8 sets is a partially symmetric test set the we call *3-mirror-symmetric*.

The identification of the opposite faces reduces the number of faces to 3, and makes it impossible to distinguish *left* and *right*, *top* and *bottom*, and *front* and *back*. Thus we can

identify cycles which are the reverse of each other, i.e. 1=2, 3=4, and 5=6. Simplifying the notation we "divide" by 2, $i \to \lceil i/2 \rceil$, $1 \leq i \leq 6$, leaving us with 3 reduced cycles denoted 1, 2, and 3. This reduces 3.6 in the following way:

$$1 \longrightarrow 2-3-2-3$$

$$2 \longrightarrow 1-3-1-3$$

$$3 \longrightarrow 1-2-1-2.$$

Similarly Figure 3.6 is reduced as follows:

$$1-2 \Rightarrow 3-1; \qquad 1-3 \Rightarrow 2-1$$

$$2-1 \Rightarrow 3-2; \qquad 2-3 \Rightarrow 1-2$$

$$3-1 \Rightarrow 2-3; \qquad 3-2 \Rightarrow 1-3.$$

Using the initial segments $A = 2\ldots$ and $B = 12\ldots$, and repeating the arguments above with the given reductions, we have

$$A = 232321212\ldots,$$

which implies

$$B = 131312321313121213131323231313231313\ldots$$

Thus again by the same logic, simply knowing the partial information $A = \ldots 321\ldots$ anywhere in the middle of $A$ guarantees the corresponding sequence
$B = \ldots 313132323\ldots$ in the middle of $B$, so we say that

$$313, 132, 323, 232, 323$$

are the new reduced consecutive triples generated (in B) by the triple 321 in the middle (of A).

Similarly, the new triples generated by 232 in the middle are

$$213, 131, 313, 131.$$

So, 321 generates 5 new triples with repetition, and 232 generates 4 new triples, again with repetition. These define the following recurrences

$$x_{n+1}(321) = x_n(313) + x_n(132) + x_n(323) + x_n(232) + x_n(323) =$$

$$= x_n(313) + x_n(132) + 2x_n(323) + x_n(232) \tag{5.1}$$

and

$$x_{n+1}(232) = x_n(213) + x_n(131) + x_n(313) + x_n(131) =$$

$$= x_n(213) + 2x_n(131) + x_n(313). \tag{5.2}$$

With this reduction, we have reduced the number of triples to $3 \cdot 2 \cdot 2 = 12$, and the eigenvalues of a $12 \times 12$ matrix can be solved without much technical difficulty. So all we need is to find 10 more equations like 5.1 and 5.2, which we can determine through brute force examination. Repeating the process above for all new triples gives the following equations (equations 5.1

and 5.2 repeated for clarity):

$$
\begin{aligned}
x_{n+1}(121) &= x_n(132) + x_n(232) + 2x_n(323) \\
x_{n+1}(123) &= x_n(131) + x_n(312) + x_n(212) + 2x_n(121) \\
x_{n+1}(131) &= x_n(123) + x_n(323) + 2x_n(232) \\
x_{n+1}(132) &= x_n(121) + x_n(213) + x_n(313) + 2x_n(131) \\
x_{n+1}(212) &= x_n(231) + 2x_n(313) + x_n(131) \\
x_{n+1}(213) &= x_n(232) + x_n(321) + 2x_n(212) + x_n(121) \\
x_{n+1}(231) &= 2x_n(232) + x_n(323) + x_n(212) + x_n(123) \\
x_{n+1}(232) &= x_n(213) + 2x_n(131) + x_n(313) \\
x_{n+1}(312) &= x_n(131) + 2x_n(313) + x_n(231) + 2x_n(323) \\
x_{n+1}(313) &= x_n(321) + 2x_n(212) + x_n(121) \\
x_{n+1}(321) &= x_n(313) + x_n(132) + 2x_n(323) + x_n(232) \\
x_{n+1}(323) &= x_n(312) + 2x_n(121) + x_n(212)
\end{aligned}
\tag{5.3}
$$

To simplify the notation in 5.3, we replace the triples in lexicographic order

(i.e. $121, 123, \ldots, 323$) with the corresponding positive integers $1, 2, \ldots, 12$ in increasing order,

thus yielding:

$$
\begin{aligned}
x_{n+1}(1) &= x_n(4)+x_n(8)+2x_n(12)\\
x_{n+1}(2) &= 2x_n(1)+x_n(4)+x_n(5)+x_n(9)\\
x_{n+1}(3) &= x_n(2)+2x_n(8)+x_n(12)\\
x_{n+1}(4) &= x_n(1)+2x_n(3)+x_n(6)+x_n(10)\\
x_{n+1}(5) &= x_n(3)+x_n(7)+2x_n(10)\\
x_{n+1}(6) &= x_n(1)+2x_n(5)+x_n(8)+x_n(11)\\
x_{n+1}(7) &= x_n(2)+x_n(5)+2x_n(8)+x_n(12)\\
x_{n+1}(8) &= 2x_n(3)+x_n(6)+x_n(10)\\
x_{n+1}(9) &= x_n(3)+x_n(7)+2x_n(10)+2x_n(12)\\
x_{n+1}(10) &= x_n(1)+2x_n(5)+x_n(11)\\
x_{n+1}(11) &= x_n(5)+x_n(8)+x_n(10)+2x_n(12)\\
x_{n+1}(12) &= 2x_n(1)+x_n(5)+x_n(9).
\end{aligned}
\tag{5.4}
$$

We next rewrite 5.4 in matrix form:

$$
M = \begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2\\
2 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 1\\
1 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0\\
1 & 0 & 0 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 & 0\\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 & 0 & 1\\
0 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 1\\
1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 2\\
2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0
\end{pmatrix}.
\tag{5.5}
$$

Then 5.4 is equivalent to $\mathbf{x}_{n+1} = M\mathbf{x}_n$ where $\mathbf{x}_j = (x_j(1), x_j(2), \ldots, x_j(12))$ is a column vector. Using any program to calculate the eigenvalues (e.g. Mathematica, Maple, online calculators) it is easy to determine the eigenvalues of matrix $M$. The results of doing so are as follows (in order of descending absolute value): $\lambda_1 = 2+\sqrt{5}$ (as we expected), $\lambda_2 = \lambda_3 = -(1+\sqrt{2})$, $\lambda_4 = \lambda_5 = \lambda_6 = 1$, $\lambda_7 = \lambda_8 = \lambda_9 = -1$, $\lambda_{10} = \lambda_{11} = \sqrt{2}-1$, $\lambda_{12} = -(\sqrt{5}-2)$. Therefore, the coordinates of

$$\mathbf{x}_n = M\mathbf{x}_{n-1} = M^n\mathbf{x}_0$$

have the generic form

$$c_1\lambda_1^n + c_2(-1)^n|\lambda_2|^n + c_3 n(-1)^n|\lambda_2|^n +$$

$$+c_4 + nc_5 + n^2c_6 + c_7(-1)^n + c_8 n(-1)^n + c_9 n^2(-1)^n + c_{10}\lambda_{10}^n + c_{11}n\lambda_{10}^n + c_{12}\lambda_{12}^n. \quad (5.6)$$

This follows from the Jordan normal form of matrices and how the powers of Jordan normal forms look. In 5.6, the first line represents the dominating part, as the base of the exponentials have absolute value greater than 1, and the second line represents the negligible part, as the base of the exponents have absolute value less than or equal to 1.

As

$$|\lambda_2| = 1+\sqrt{2} > \sqrt{\lambda_1} = \sqrt{2+\sqrt{5}}, \quad (5.7)$$

this implies that the fluctuations of size

$$c_2(-1)^n|\lambda_2|^n + c_3 n(-1)^n|\lambda_2|^n$$

around the main term $c_1\lambda_1^n$, which are substantially larger than square root size fluctuations. If we denote the main term $X$, then the fluctuations have the order

$$X^{c_0} \text{ with } c_0 = \frac{\log(1+\sqrt{2})}{\log(2+\sqrt{5})}. \quad (5.8)$$

There is one improvement to 5.8 that we need that keeps it from being a complete result. The alternating process of generating cycle sequences back and forth between $A$ and $B$ generates sequences of length $2, 9, 38, \ldots$ and so on, increasing roughly by a factor of $2+\sqrt{5}$ each

time. This means that 5.8 holds for cycle sequences of these lengths, but this will only be an exponentially sparse set of numbers. To prove good quantitative results about combinatorial uniformity of face-crossing we need information about every integer length $N$, not just exponentially sparse "special length numbers". (Note that because 5.8 says maximum fluctuation size at these special length numbers has exponent greater than $\sqrt{2+\sqrt{5}}$, the overall maximum must as well, and thus our conclusion of this being a new type of pseudo-random chaos still holds.)

Additionally, another improvement to our quantitative results it to look beyond "combinatorial uniformity". This notion of "face-crossing sequence" is a bit contrived, and dissimilar from our other results in that it does not refer to 2-dimensional Lebesgue uniformity on the surface of the cube—that is, uniformity for a given test set on the surface of the cube. These two seemingly unrelated improvements actually have similar resolutions, essentially decomposing line lengths into relatively few "special length numbers". In the next chapter we strengthen the result in 5.8 by resolving these two improvements.

# CHAPTER 6

# QUANTITATIVE RESULTS: EDGE LENGTH UNIFORMITY

We look now at the continued fraction expansion of

$$\alpha = \sqrt{5} - 2 = \frac{1}{4 + \frac{1}{4 + \frac{1}{4 + \ldots}}}.$$

It is quite simple, and so we can easily find the explicit form of the convergents $p_i/q_i$ of $\alpha$. We have that

$$\frac{p_1}{q_1} = \frac{1}{4}, \ \frac{p_2}{q_2} = \frac{4}{17}, \ \frac{p_3}{q_3} = \frac{17}{72},$$

and in general we get the linear recurrence relation $p_{i+2} = 4p_{i+1} + p_i$. The corresponding characteristic polynomial is $x^2 - 4x - 1 = 0$ with roots $x_1 = 2 + \sqrt{5} = 1/\alpha$ and $x_2 = 2 - \sqrt{5} = -\alpha$. Using these two roots, and the initial condition $p_1 = 1, p_2 = 4$, we get the explicit formula

$$q_{i-1} = p_i = \frac{1}{2\sqrt{5}} \left( (2+\sqrt{5})^i - (2-\sqrt{5})^i \right) = \frac{1}{2\sqrt{5}} \left( \alpha^{-1} - (-\alpha)^i \right), \ \ i \geq 2. \tag{6.1}$$

Further evaluating 6.1 yields

$$q_i\alpha - p_i = \frac{1}{2\sqrt{5}} \left( (\alpha)^{-i-1} - (-\alpha)^{i+1} \right) \alpha - \frac{1}{2\sqrt{5}} \left( \alpha^{-1} - (-\alpha)^i \right), =$$

$$= \frac{1}{2\sqrt{5}}\left((-\alpha)^i - (-\alpha)^{i+2}\right) = \frac{1+\alpha^2}{2\sqrt{5}}(-\alpha)^i = (-1)^i\alpha^{i+1}, \;\; i \geq 1. \tag{6.2}$$

The key to proving the two upgrades mentioned at the end of Chapter 5 is a decomposition into logarithmically few "special numbers", where these special numbers are the irrational powers $\alpha^{2j-1}, \; j \geq 1$. These numbers come from the lengths of the sequences of cycles generated for $A$ and $B$ as in 3.7 and 3.8, while taking the second power because the sequence in 6.2 alternates. This creates what is basically an analog of the decimal expansion of the real numbers between 0 and 1, in that every real number $0 < y < 1$ can be written as

$$y = b_1(\sqrt{5}-2) + b_2(\sqrt{5}-2)^3 + b_3(\sqrt{5}-2)^5 + b_4(\sqrt{5}-2)^7 + \ldots, \tag{6.3}$$

where $b_1 \in \{0,1,2,3,4\}$, $b_i \in \{0,1,2,\ldots,16\}$, $i \geq 2$ (here the 16 comes from the fact that $16 < (\sqrt{5}-2)^2 < 17$).

We next consider $A(t), \; 0 < t < T_k^*$, where $T_k^*$ are special numbers, which we will call a *special length* initial segment. We know that a special length sequence $T_k^*$ is exponentially sparse: it grows basically like $(2+\sqrt{5})^{2k}$. Consider an arbitrary but fixed triple of consecutive cycles, say, 162. Then $A(t), \; 0 < t < T_k^*$ generates

$$c(162)T_k^* + O((T_k^*)^{\gamma_0}) \tag{6.4}$$

copies of the given triple 162, where $c(162)$ is a constant dependent only on the specific triple 162. In general, we get equations similar to 10.4 for all consecutive triples, with the natural change that the constant term $c(.)$ depends only on the given triple. Here we have that $\gamma_0$ is in the interval

$$1/2 < c_0 \leq \gamma_0 < 1, \tag{6.5}$$

where $c_0$ is as defined in 5.8, and $\gamma_0$ is an absolute constant that depends only the second largest eigenvalue of the transition matrix given in 5.5.

The simplest special length sequence is

$$T_k^* = q_k\sqrt{1+\alpha^2}, \;\; k \geq 1, \tag{6.6}$$

and the corresponding cycle-counting sequence is $p_k$ (where once again $p_i/q_i$ are the convergents of $\alpha$). The initial segment $A(t),\ 0 < t \leq T_k^* = q_k\sqrt{1+\alpha^2}$ has several $\theta$-edge crossings; consider the crossing closest to a vertex. We know this will occur when a multiple of $\alpha$ is very close to a whole number, which due to the nature of convergents will occur at the endpoint $A\left(q_k\sqrt{1+\alpha^2}\right)$ with distance $||q_k\alpha|| = \alpha^{k+1}$, where $||.||$ denotes the distance of a real number to the nearest integer. Consider the point $(0,\alpha^{k+1})$ on the boundary of the unit square $[0,1]^2$, and start a half-line from this point going down and right with slope $\alpha$ relative to the vertical. This line intersects the low horizontal edge of the unit square at the point $(\alpha^{k_2},0)$. Note that this point would be a $\theta$-edge crossing of $B$, as it is the meta-line of $A$, which in turn tells us that

$$\alpha||q_k\alpha|| = \alpha^{k+2} = ||q_{k+1}\alpha||. \tag{6.7}$$

We see that $q_{k+1}\sqrt{1+\alpha^2}$ is the next element in the special length sequence that started with $T_j^*,\ 1 \leq j \leq k$. Note that the sequence $q_j,\ j \geq 1$ of convergence denominators of $\alpha$ represents the local minima of the best rational approximations of $\alpha$—that is, it satisfies

$$\min_{1\leq m\leq q_{j+1}} ||m\alpha|| = ||q_j\alpha||. \tag{6.8}$$

We consider a directed edge, $\vec{\mathscr{E}}$, of the cube, and recall that the cube line $A(t)$ will hit $\vec{\mathscr{E}}$ with $\alpha$, $\theta$-edge crossings, and with slope $1/\alpha$, $\theta^{\perp}$-edge crossings. Indeed we recall from Chapter 3 that the crossings with slope $\alpha$ of $A(t)$ will be the crossings with slope $1/\alpha$ of $B(t)$, and similarly the crossings with slope $\alpha$ of $B(t)$ are the $1/\alpha$ crossings of $A(t)$. We will use the special length sequence in 6.6 to make this more explicit in the form of a lemma. We first note however that every integer $n \leq 1$ can be written as

$$n = \sum_{i=0}^{r} b_i q_i \text{ with } b_r \in \{1,2,3,4\},\ b_i \in \{0,1,2,3,4\},\ 0 \leq i < r, \tag{6.9}$$

which can be made to be unique by imposing the restriction

$$b_j = 4 \text{ only if } b_{j-1} = 0,\ j \leq r. \tag{6.10}$$

The restriction in 6.10 follows from the recurrence formula $q_{j+1} = 4q_j + q_{j-1}$. Together 6.9 and 6.10 form a complete and unique characterization of the positive integers $\geq 1$, which we will call the $\alpha$-*representation* of $n$.

**Lemma 6.** *Let $n \geq 1$ be an integer, and consider its $\alpha$-representation*

$$n = \sum_{i=0}^{r} b_i q_i \text{ with } b_r \in \{1,2,3,4\},\ b_i \in \{0,1,2,3,4\},\ 0 \leq i < r.$$

*Let*

$$N(n) = \sum_{i=0}^{r} b_i q_{i+1} \text{ with } b_r \in \{1,2,3,4\},\ b_i \in \{0,1,2,3,4\},\ 0 \leq i < r,$$

*i.e. $N(n)$ is obtained by replacing the $q_i$ in the $\alpha$-representation of $n$ with $q_{i+1}$ for every $0 \leq i < r$, which is essentially a multiplication by $1/\alpha$. The edge crossings with slope $\alpha$ in $A(t)$, $0 < t < \sqrt{1+\alpha^2}N(n)$ are the edge crossings with slope $1/\alpha$ of $B(t)$, $0 < t < \sqrt{1+\alpha^2}n$; and similarly the edge crossings with slope $\alpha$ in $B(t)$, $0 < t < \sqrt{1+\alpha^2}N(n)$ are the edge crossings with slope $1/\alpha$ of $A(t)$, $0 < t < \sqrt{1+\alpha^2}n$.*

See Chapter 3 for justification. □

The following lemma is about the number of times $A(t)$ crosses $\vec{\mathscr{E}}$ with slope $\alpha$ and with slope $1/\alpha$.

**Lemma 7.** *Let $\vec{\mathscr{E}}$ be any of the 24 directed edges of the cube. Then $A(t),\ 0 < t < T$ has*

$$\frac{T}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0}\log T) \tag{6.11}$$

*$\vec{\mathscr{E}}$ crossings with slope $1/\alpha$, and has*

$$\alpha\frac{T}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0}\log T) \tag{6.12}$$

*$\vec{\mathscr{E}}$ crossings with slope $\alpha$.*

**Proof.** It suffices to prove Lemma 7 for the sub-sequences $T = \sqrt{1+\alpha^2}n$, where $n \geq 1$ is an integer. Consider the $\alpha$-representation of $n$ as described in 6.9. We prove this lemma by

induction on the "digit-sum"

$$D(n) = \sum_{i=0}^{r} b_i. \tag{6.13}$$

If $D(n) = 1$, then $b_r = 1$, and $b_i = 0$ for all $0 \leq i < r$, and Lemma 7 follows from 6.4. Indeed we use the fact that each cycle transition on a cube-line happens on a specific directed edge (see Figure 3.6). It then follows from 6.4 and our "soft" qualitative uniformity of $A$ in Theorem 4 that $A(t),\ 0 < t < T_k^*$ has

$$\frac{T_k^*}{24\sqrt{1+\alpha^2}} + \text{Error}_1 \tag{6.14}$$

$\vec{\mathscr{E}}$ crossings with slope $1/\alpha$, and has

$$\alpha \frac{T_k^*}{24\sqrt{1+\alpha^2}} + \text{Error}_2 \tag{6.15}$$

$\vec{\mathscr{E}}$ crossings with slope $\alpha$, where

$$|\text{Error}_1| \leq C_1 (T_k^*)^{\gamma_0} \text{ and } |\text{Error}_2| \leq C_1 (T_k^*)^{\gamma_0}. \tag{6.16}$$

Theorem 4 explains why there must be the same factor of 24 in the denominator of both main terms, regardless of which directed edge $\vec{\mathscr{E}}$ is chosen (here the factor of 24 comes from the total number of directed edges).

If $D(n) > 1$ then consider the initial segment $A(t),\ 0 < t < T_k^* = \sqrt{1+\alpha^2} q_r$ of $A(t),\ 0 < t < T$. The previous argument proves that this initial segment has crossings of slope $\alpha$ and $1/\alpha$ as described in 6.14-10.15.

Moreover, as in 6.7, the cross-point nearest to a vertex, say, $V$ of the cube occurs at $A\left(q_k\sqrt{1+\alpha^2}\right)$ with distance $||q_k\alpha|| = \alpha^{k+1}$. Vertex $V$ is the corner of 3 faces of the cube. There is one face $F$ where during $T_k^* < t < T_k^* + 1$ spends, by far, the longest amount of time, say during $T_k^* + \varepsilon < t < T_k^* + 1$. Let $A'(t)$ denote the cube line with initial slope $\alpha$ which starts at the vertex $V$, i.e. $A'(0) = V$. Then the two cube lines $A'(t),\ 0 < t < 1$ and $A(t)$ $T_k^* + \varepsilon < t < T_k^* + 1$ are very close parallel line segments on face $F$. It follows from the local minimum property 6.8 that $A'(t),\ 0 < t < T - T_k^*$ and $A(t)$ $T_k^* < t < T$ will remain very close parallel orbits, crossing exactly the same edges and the same faces (6.8 tells us no integers will lie between the lines, and thus they will not be separated by any singularities).

Our induction hypothesis tells us that $A'(t),\ 0 < t < T - T_k^*$ has

$$\frac{T'}{24\sqrt{1+\alpha^2}} + \text{Error}_3$$

$\vec{\mathscr{E}}$ crossings with slope $1/\alpha$, and has

$$\alpha\frac{T'}{24\sqrt{1+\alpha^2}} + \text{Error}_4$$

$\vec{\mathscr{E}}$ crossings with slope $\alpha$, where

$$|\text{Error}_3| \leq \left(b_r - 1 + \sum_{i=0}^{r-1} b_i\right) C_1 (T')^{\gamma_0} \text{ and } |\text{Error}_2| \leq \left(b_r - 1 + \sum_{i=0}^{r-1} b_i\right) C_1 (T')^{\gamma_0}.$$

Therefore $A(t)\ T_k^* < t < T$ will have exactly the same number of crossings, i.e. it will have

$$\frac{T'}{24\sqrt{1+\alpha^2}} + \text{Error}_3 \tag{6.17}$$

$\vec{\mathscr{E}}$ crossings with slope $1/\alpha$, and will have

$$\alpha\frac{T'}{24\sqrt{1+\alpha^2}} + \text{Error}_4 \tag{6.18}$$

$\vec{\mathscr{E}}$ crossings with slope $\alpha$, where $T' = T - T_k^*$ and

$$|\text{Error}_3| \leq \left(b_r - 1 + \sum_{i=0}^{r-1} b_i\right) C_1 (T')^{\gamma_0} \text{ and } |\text{Error}_2| \leq \left(b_r - 1 + \sum_{i=0}^{r-1} b_i\right) C_1 (T')^{\gamma_0}. \tag{6.19}$$

Combining 6.14-6.19 we get that $A(t)\ 0 < t < T_k^* + T'$ has

$$\frac{T_k^* + T'}{24\sqrt{1+\alpha^2}} + \text{Error}_5 \tag{6.20}$$

$\vec{\mathscr{E}}$ crossings with slope $1/\alpha$, and has

$$\alpha\frac{T_k^* + T'}{24\sqrt{1+\alpha^2}} + \text{Error}_6 \tag{6.21}$$

$\vec{\mathscr{E}}$ crossings with slope $\alpha$, where

$$|\text{Error}_5| \leq |\text{Error}_1| + |\text{Error}_3| \leq D(n)C_1(T)^{\gamma_0}$$

$$|\text{Error}_6| \leq |\text{Error}_2| + |\text{Error}_4| \leq D(n)C_2(T)^{\gamma_0}. \tag{6.22}$$

This completes the proof of Lemma 7. □

We next include the following lemma, which is a simple corollary to Lemma 7 and allows us to apply the results to cube lines that do not start at a vertex.

**Lemma 8.** *Let $\vec{\mathscr{E}}$ be any of the 24 directed edges of the cube. Let $L_\alpha(t),\ 0 < t < T$ be an arbitrary cube line segment of slope $\alpha = \sqrt{5} - 2$. (Here $L_\alpha(0)$ is not necessarily a vertex, but as always we assume $L_\alpha(t),\ 0 < t < T$ does not hit a vertex.) Then $L_\alpha(t),\ 0 < t < T$ has*

$$\frac{T}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0}\log T)$$

*$\vec{\mathscr{E}}$ crossings with slope $1/\alpha$, and has*

$$\alpha\frac{T}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0}\log T)$$

*$\vec{\mathscr{E}}$ crossings with slope $\alpha$.*

The justification of this lemma follows the same argument as the inductive step in the proof of Lemma 7. We simply take the edge crossing of $L_\alpha(t)$ which has minimal distance to a vertex, say, $V$ of the cube. We then consider the cube line $A'(t)$ with slope $\alpha$ and initial point $A'(0) = V$ which runs parallel to $L_\alpha(t)$ (in the forward direction). Applying Lemma 7 to $A'(t)$, and using the same argument as in the proof of Lemma 7, this implies the result also holds for $L_\alpha(t)$. Doing this for both directions, i.e. considering the cube line $A''(t)$ with slope $-\alpha$ and initial point $A''(0) = V$ which runs parallel to $L_\alpha(t)$ in the negative direction, completes the proof. □

For each directed edge $\vec{\mathscr{E}}$ we use the usual induced parametrization $\vec{\mathscr{E}} = \vec{\mathscr{E}}[0,1]$. Consider an arbitrary edge-crossing of $A(t),\ 0 < t < T$ on $\vec{\mathscr{E}}$ with slope $\alpha$. Now move backwards

from this point along the flow with slope $1/\alpha$. We eventually cross an edge perpendicular to $\vec{\mathscr{E}}$, call it $\vec{\mathscr{E}}_1$. This crossing has slope $1/\alpha$ and is on $\vec{\mathscr{E}}_1[1-\alpha,1]$. We continue moving backwards along the flow with slope $1/\alpha$ to get to an edge $\vec{\mathscr{E}}_2$ which is parallel to $\vec{\mathscr{E}}_1$, and again we have an edge crossing this time on $\vec{\mathscr{E}}_2[1-2\alpha,1-\alpha]$. We continue this process, next getting a crossing on $\vec{\mathscr{E}}_3[1-3\alpha,1-2\alpha]$, and so on. This yields the general form

$$\frac{1}{\alpha}\text{-slope } \vec{\mathscr{E}}_{k+1}[\{1-(k+1)\alpha\},\{1-k\alpha\}]\text{-crossing} \tag{6.23}$$

for some directed edge $\vec{\mathscr{E}}_{k+1}$, where as usual $\{.\}$ represents the fractional part of a real number. Note the caveat that for some integer values of $k$, $\{1-(k+1)\alpha\} > \{1-k\alpha\}$, in which case their order in 6.23 is reversed.

We let $0 \le k \le T$, and claim that $A(t),\ 0 < t < T$ has

$$\frac{T}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0}\log T)$$

$$\frac{1}{\alpha}\text{-slope } \vec{\mathscr{E}}_{k+1}[\{1-(k+1)\alpha\},\{1-k\alpha\}]\text{-crossings} \tag{6.24}$$

for any directed edge $\vec{\mathscr{E}}$, noting that $[\{1-(k+1)\alpha\},\{1-k\alpha\}]$ has length $\alpha$. To prove 6.24 we simply first apply Lemma 7 to the long segment
$A(t),\ 0 < t < T + k\sqrt{1+\alpha^2}$, then we apply Lemma 8 to the shorter sub-segment $L_\alpha(t) = A(t+T),\ 0 < t < T + k\sqrt{1+\alpha^2}$, and then finally take the difference of the two.

For notational convenience we merge together the cases of possible intervals in 6.23 to a single notation. We write

$$I_{1,k} = [\{1-(k+1)\alpha\},\{1-k\alpha\}] \text{ if } \{1-(k+1)\alpha\} < \{1-k\alpha\} \text{ and}$$

$$I_{1,k} = [\{1-k\alpha\},\{1-(k+1)\alpha\}] \text{ if } \{1-(k+1)\alpha\} > \{1-k\alpha\}. \tag{6.25}$$

We note that $I_{1,k}$ is a sub-interval of the one-dimensional unit torus $[0,1)$ of length $\alpha$. The intervals $I_{1,k}$ $0 \le k \le T$ are "dense" in the unit torus $[0,1)$ in the sense that every sub-interval $J \subset [0,1)$ of the unit torus with length $10/T$ contains two points $P_1, P_2$ with the property that $P_1$

is an endpoint of an interval $I_{1,k_1}$, $0 \leq k_1 \leq T$ such that $I_{1,k_1}$ goes from $P_1$ to its other endpoint in positive orientation on the unit torus, and $P_2$ similarly is an endpoint of an interval $I_{1,k_2}$, $0 \leq k_2 \leq T$ such that $I_{1,k_2}$ goes from $P_2$ to its other endpoint, only this time in the negative orientation on the unit torus. We call this property *quantitative density*.

Applying 6.2 we get

$$\left| i\alpha - \frac{ip_j}{q_j} \right| = \frac{i\alpha^{j+1}}{q_j} < \alpha^{j+1} < \frac{1}{2q_j} \text{ for all } 0 \leq i < q_j, \tag{6.26}$$

with the last part following from 6.1. As $q_j$ and $p_j$ are relatively prime, 6.26 implies that for every interval $U_r$—defined as

$$U_r = \left[ \frac{r-(1/2)}{q_j}, \frac{r+(1/2)}{q_j} \right],\ 1 \leq r \leq q_j - 1$$

and

$$U_{q_j} = \left[ 0, \frac{1/2}{q_j} \right] \cup \left[ \frac{q_j-(1/2)}{q_j}, 1 \right],$$

—contains exactly one element of the set $\{\{i\alpha\} :\ 0 \leq i < q_j\}$ of fractional parts. Finally we note

$$q_j < T \leq q_{j+1} \text{ with } q_j \geq T/5.$$

This proves the quantitative density of the intervals $I_{1,k}$ $0 \leq k \leq T$.

Now we use Lemma 6; it follows from this that the $A(t),\ 0 < t < T$, $\vec{\mathscr{E}} I_{1,k}$-crossings with slope $1/\alpha$ are the $B(t), 0 < t < T/\alpha$, $\vec{\mathscr{E}} I_{1,k}$-crossings with slope $\alpha$. We consider an arbitrary $\vec{\mathscr{E}} I_{1,k}$-crossing of $B(t), 0 < t < T/\alpha$ and repeat the process as in 6.23, and travel backwards along the $1/\alpha - B-$flow until we cross a perpendicular edge, say, $\vec{\mathscr{E}}_1$. This gives us a $\vec{\mathscr{E}} I_{2,k} - B$-crossing with slope $1/\alpha$, where $I_{2,k}$ is defined as

$$I_{1,k} = [v_1, v_2], \quad I_{2,k} = [\beta_1, \beta_2],$$

$$\beta_1 = 1 - \alpha v_2, \quad \beta_2 = 1 - \alpha v_1. \tag{6.27}$$

It then follows from 6.24 that, for every integer $0 \leq k \leq T$, $B(t), 0 < t < T_1 = T/\alpha$ has

$$\alpha \frac{T}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0} \log T) =$$

$$= \alpha^2 \frac{T}{24\sqrt{1+\alpha^2}} + O((\alpha T_1)^{\gamma_0} \log T_1) \tag{6.28}$$

$\vec{\mathscr{E}} I_{2,k}$-$B$-crossings with slope $1/\alpha$ for any directed edge $\vec{\mathscr{E}}$. Note that $I_{2,k}$ has length $\alpha^2$.

We note that the set of intervals $\vec{\mathscr{E}} I_{2,k},\ 0 \leq k \leq T$ are quantitatively dense in the sub-interval $[1-\alpha, 1)$ of the unit torus; that is, every sub-interval $J \subset [1-\alpha, 1)$ of the unit torus with length $10\alpha/T = 10/T_1$ contains two points $P_1, P_2$ with the property that $P_1$ is an endpoint of an interval $I_{2,k_1}$, $0 \leq k_1 \leq T$ such that $I_{2,k_1}$ goes from $P_1$ to its other endpoint in positive orientation on the unit torus, and $P_2$ is an endpoint of an interval $I_{2,k_2}$, $0 \leq k_2 \leq T$ such that $I_{2,k_2}$ goes from $P_2$ to its other endpoint in the negative orientation on the unit torus.

This quantitative density can be extended to the entire unit torus $[0,1)$ by 4 extra rotations of size $\alpha$. We can continue the process used to obtain 6.27 in the same way we did with 6.23, that is, continuing to travel backwards along the $1/\alpha$-$B$-flow to get more edge crossings of slope $1/\alpha$. By the same logic as above, the set of intervals generated from repeating the process above on the nextfour edge crossings will be quantitatively dense on $[1-2\alpha, 1-\alpha], [1-3\alpha, 1-2\alpha]$, $[1-4\alpha, 1-3\alpha], [1-5\alpha, 1-4\alpha]$. As $5 > 4\alpha$, the set of all these sub-intervals $I_{2_i,k}$ for all 5 rotations is quantitatively dense over the unit torus $[0,1)$ (here the $i$ in $I_{2_i,k}$ denotes the different rotations, $i = 0,1,2,3,4$).

We refer to the sub-intervals $I_{1,k},\ 0 \leq k \leq T$ as *special sub-intervals* of type $(1;T)$, and refer to the sub-intervals $I_{2_i,k},\ 0 \leq k \leq T,\ i = 0,1,2,3,4$ as *special sub-intervals* of type $(2;T_1)$ with $T_1 = T/\alpha$.

We continue to repeat this process, converting the edge crossings with slope $1/\alpha$ of $B(t)$ into edge crossings with slope $\alpha$ of $A(t)$, which are then used to generate edge crossings with slope $1/\alpha$ of $A(t)$, which are converted back to edge crossings with slope $\alpha$ of $B(t)$, and so on, repeating the process above to generate a complete set of *special sub-intervals* of type $(n;T_{n-1})$ for all positive integers $n \geq 1$.

Consider the cube line segment $A(t),\ 0 < t < N$; let $U \subset [0,1)$ be an arbitrary sub-interval

with $U=[u_1,u_2], 0<u_1<u_2<1$. If, say, $|U|=u_2-u_1\geq\alpha$, then we chose a special sub-interval of type $(1;N)$, $I_1=[x_1,y_1]$ that contains $u_1$. By our quantitative density property, we can assume $|u_1-x_1|\leq 10/N$. If, say, $|u_2-y_1|\geq\alpha$ then we chose a special sub-interval of type $(1;N)$, $I_1=[x_2,y_2]$ that contains $y_1$. By our quantitative density property, we can assume $|y_2-x_2|\leq 10/N$. Now suppose that, say, $\alpha^3\leq u_2-y_2<\alpha$; we then chose a special sub-interval of type $(3;N)$, $I_3=[x_3,y_3]$ that contains $y_2$. By our quantitative density property, we can assume $|x_3-y_2|\leq 10/N$. If, say, $|u_2-y_3|\geq\alpha^3$, then we chose a special sub-interval of type $(3;N)$, $I_4=[x_4,y_4]$ that contains $y_3$. By our quantitative density property, we can assume $|x_4-y_3|\leq 10/N$, and so on.

This example demonstrates an economic covering of the given interval $U$ with "few" special sub-intervals of odd types $(1;N),(3;N),\ldots$ (we use only odd types due to the alternations between $A(t)$ and $B(t)$). This is basically an analog of the decomposition in 6.3. We call it "economic" because, thanks to the quantitative density property, this covering has very small overlaps. We stop at type $(2\ell+1;N)$, where $\ell$ is the smallest integer such that

$$\alpha^{2\ell+1}N<N^{\gamma_0}, \tag{6.29}$$

which implies that $\ell=O(\log N)$. We can cover the last bit of $U$ which is still uncovered (the part near $u_2$) in a trivial way with a couple of special sub-intervals of type $(2\ell+1;N)$ (these last intervals will be small enough and few enough that their overlapping is negligible). Thus we have a complete cover $I_1\cup I_2\cup I_3\cup\ldots$ of $U$ with very small overlappings; $U$ is *almost* partitioned by $O(\log N)$ special sub-intervals. For these special sub-intervals we can apply our 6.24 type results, and adding them up we get that $A(t),\ 0<t<N$ has at most

$$\left(\sum_{i\geq 1}|I_i|\right)\frac{N}{24\sqrt{1+\alpha^2}}+\left(\sum_{i=0}^{\infty}\alpha^{2i\gamma_0}\right)O(N^{\gamma_0}\log N)=$$

$$\left(\sum_{i\geq 1}|I_i|\right)\frac{N}{24\sqrt{1+\alpha^2}}+O(N^{\gamma_0}\log N) \tag{6.30}$$

$\vec{\mathscr{E}}U$-$A$-crossings of slope $1/\alpha$ for any directed edge $\vec{\mathscr{E}}$.

As there are $O(\log N)$ overlappings, each of which has size at most $10/N$, plus the last step

which represents an extra contribution of $O(N^{-1+\gamma_0})$(see 6.29), we can bound the total overlap to obtain

$$\sum_{i\geq 1} |I_i| \leq |U| + O(\log N/N) + O(N^{-1+\gamma_0}). \tag{6.31}$$

Combining 6.30 with the bound given by 6.31 we get that there are at most

$$|U|\frac{N}{24\sqrt{1+\alpha^2}} + O(N^{\gamma_0}\log N) \tag{6.32}$$

$\vec{e}$$U$-$A$-crossings of slope $1/\alpha$ in $A(t),\ 0 < t < N$ for any sub-interval $U$ of any directed edge $\vec{e}$.

This represents an outside approximation—that is, the special sub-intervals overlap each other and cover (with excess) $U$. We can repeat a similar argument for an inside approximation, by finding disjoint special sub-intervals which are contained within $U$, and which remaining part of $U$ that is not covered can be bound as very small given the quantitative density of the special sub-intervals, just as the overlappings were in 6.31. Repeating the above argument (with the straightforward modifications) yields an analog of 6.32—there are **at least**

$$|U|\frac{N}{24\sqrt{1+\alpha^2}} + O(N^{\gamma_0}\log N) \tag{6.33}$$

$\vec{e}$$U$-$A$-crossings of slope $1/\alpha$ in $A(t),\ 0 < t < N$ for any sub-interval $U$ of any directed edge $\vec{e}$. By combining the two bounds in 6.32 and 6.33 we obtain the following lemma.

**Lemma 9.** *Let $\vec{e}$ be any of the 24 directed edges of the cube, and let $0 < y < 1$ be any real number. Then $A(t),\ 0 < t < T$ has*

$$\frac{yT}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0}\log T)$$

$\vec{e}[0,y]$*-crossings of slope $1/\alpha$, and*

$$\frac{\alpha yT}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0}\log T)$$

$\vec{e}[0,y]$*-crossings of slope $\alpha$.* □

We strengthen Lemma 9 in the same way we strengthened Lemma 7 to get

Lemma 8.

**Theorem 5.** *Let $\vec{\mathscr{E}}$ be any of the 24 directed edges of the cube. Let $L_\alpha(t),\ 0<t<T$ be an arbitrary cube line segment of slope $\alpha=\sqrt{5}-2$. ($L_\alpha(0)$ is not necessarily a vertex, but as always we assume $L_\alpha(t),\ 0<t<T$ does not hit a vertex.) Then $L_\alpha(t),\ 0<t<T$ has*

$$\frac{yT}{24\sqrt{1+\alpha^2}}+O(T^{\gamma_0}\log T)$$

*$\vec{\mathscr{E}}[0,y]$-crossings of slope $1/\alpha$, and*

$$\frac{\alpha yT}{24\sqrt{1+\alpha^2}}+O(T^{\gamma_0}\log T)$$

*$\vec{\mathscr{E}}[0,y]$-crossings of slope $\alpha$. In both cases the constant $O(.)$ is effectively computable.*

**Proof.** The proof to this theorem is a direct parallel to the proof of Lemma 8. We take the edge crossing of $A(t),\ 0<t<T$ which has minimal distance to a vertex $V$ of the cube, and construct two cube lines, $A'(t)$ and $A''(t)$ which have initial point $V=A'(0)=A''(0)$, and run parallel to $A(t)$ in the forwards and backwards directions. As Lemma 9 applies to $A'(t)$ and $A''(t)$, and there are no singularities between $A(t)$ and either $A'(t)$ or $A''(t)$ on $0<t<T$, this implies Lemma 9 applies to $A(t)$ as well. □

Thus we have a quantitative result bounding the fluctuations of the edge crossings. The final part of this chapter uses this bound on the fluctuations of the edge crossings to formulate a quantitative bound on the measure-theoretic uniformity.

**Theorem 6.** *Let $L_\alpha(t),\ 0<t<T$ be an arbitrary cube line segment of slope $\alpha=\sqrt{5}-2$. Let $S$ be an arbitrary convex set on any face of the unit cube $[0,1]^3$. Then*

$$\left|\frac{1}{T}meas\{t\in[0,T]:L_\alpha(t)\in S\}-\frac{area(S)}{6}\right|=O(T^{\gamma_0-1}\log T).$$

*Here meas stands for the one-dimensional Lebesgue measure, and the implicit constant in $O(.)$ is effectively computable.*

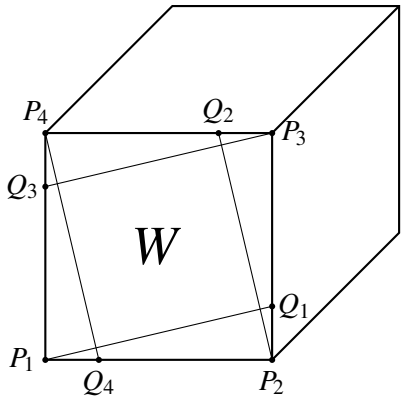


Figure 6.1: We decompose all convex test set into two types of distict convex sets: convex sets that lie within $W$ and convex sets that lie outside of $W$.

Let $S$ be a convex set on the front face of the unit cube $[0,1]^3$. Consider a tilted sub-square on the front face of the cube $(x = 1)$, defined as follows. Let the four corners of the front face be denoted $P_1 = (1,0,0),\ P_2 = (1,0,1),\ P_3 = (1,1,0),\ P_4 = (1,1,1)$, and let $Q_1 = (1,1,\alpha),\ Q_2 = (1,1-\alpha,1),\ Q_3 = (1,0,1-\alpha),\ Q_4 = (1,\alpha,0)$ be four points on the boundary of the front face. Then the line segments $P_1Q_1, P_2Q_2$,
$P_3Q_3, P_4Q_4$ divide the front face into four triangles, four trapezoids, and one tilted square, see Figure 6. We let $W$ denote this tilted square. We consider two cases.

**Case 1:** $S \subseteq W$

Let $Q = (1,y,0)$ be a point on the bottom edge of the front face with $\alpha < y < 1$. Consider the line with slope $1/\alpha$ coming from $Q$; it can be parametrized by $(1, y - z\alpha, z),\ 0 \le z \le 1$. Let $\ell(S;y)$ denote the length of the interval in which this line intersects the convex set $S$.

Defining this as a function

$$f(y) - \ell(S;y),\ \alpha \le y \le 1,$$

$f(y)$ is clearly continuous, and moreover as $S$ is convex, $f(y)$ obeys the following "increasing-decreasing" property—there exists $\alpha \le y^{(1)} \le y^{(2)} \le y^{(3)} \le 1$ such that: $f(y) = 0$ for $0 \le y \le y^{(1)}$, $f(y)$ is increasing for $y^{(1)} \le y \le y^{(2)}$, $f(y)$ is decreasing for $y^{(2)} \le y \le y^{(3)}$, and $f(y) = 0$ again for $y^{(3)} \le y \le 1$.

Let $Q^* = (1, y^*, 0)$ be a point on the bottom edge of the front face with an edge crossing of $L_\alpha(t),\ 0 < t < T$, with a slope of $1/\alpha$ directed towards the right edge of the front face. Let $Y^*(bottom; T)$ denote the set of all such $y^*$ as defined by $Q^*$—that is, all the points along the bottom edge which $L_\alpha(t)$ hits (in this particular direction and slope).

Let $g(y),\ 0 \leq y \leq 1$ denote the length of the vertical component of the $f(y)$,

$$g(y) = (1+\alpha^2)^{-1/2} f(y),\ 0 \leq y \leq 1. \tag{6.34}$$

Then we have

$$\int_0^1 g(u)du = \text{area}(S). \tag{6.35}$$

From here we look to apply Koksma's Lemma, a well known tool in uniform distribution (proof first published by Kuipers [13][15]).

**Lemma 10** (Koksma's Lemma)**.** *Let $\mathscr{X} = \{x_1, x_2, \ldots, x_n\}$ be an arbitrary n-element point set in the unit interval $[0,1)$, and let g be a function with total variation $V(g)$ on the unit interval. Then*

$$\left| \sum_{i=1}^{n} g(x_i) - n \int_0^1 g(u)du \right| \leq Discr(\mathscr{X}) V(g),$$

*where $Discr(\mathscr{X})$ denotes the discrepancy of set $X \subset [0,1)$ defined as*

$$Discr(\mathscr{X}) = \max_{0 \leq z \leq 1} \left| \sum_{x_i \leq z} 1 - nz \right|.$$

Then we can let $V(g)$ denote the total variation of $g$ (given by 6.34) on the unit interval. We clearly have

$$V(g) = (1+\alpha^2)^{-1/2} V(f) \leq 2. \tag{6.36}$$

Applying Lemma 10 with

$$\mathscr{X} = \mathscr{X}_1 = Y^*(bottom; T). \tag{6.37}$$

Theorem 5 tell us that

$$\text{Discr}(\mathscr{X}_1) = O(T^{\frac{1}{2}} \log T). \tag{6.38}$$

Then combining Lemma 10 with 6.36-6.38, we have

$$\left| \sum_{x_i \in \mathscr{X}_1} g(x_i) - |\mathscr{X}_i| \int_0^1 g(u)du \right| \leq \text{Discr}(\mathscr{X}_1) V(g) = O(T^{\gamma_0} \log T). \tag{6.39}$$

Using 6.6 and 6.36 with 6.39 we get

$$\left| \sum_{x_i \in \mathscr{X}_1} g(x_i) - \left( \frac{T}{24\sqrt{1+\alpha^2}} + O(T^{\gamma_0} \log T) \right) \text{area}(S) \right| = O(T^{\gamma_0} \log T). \tag{6.40}$$

This then implies

$$\left| \sum_{x_i \in \mathscr{X}_1} g(x_i) - \frac{T}{24\sqrt{1+\alpha^2}} \text{area}(S) \right| = O(T^{\gamma_0} \log T). \tag{6.41}$$

Here, 6.41 is about the contribution of the bottom edge of the front face. Next we replace the bottom edge with the other 3 edges to get

$$\mathscr{X}_2 = Y^*(\textit{right}; T),\ \mathscr{X}_3 = Y^*(\textit{top}; T),\ \mathscr{X}_4 = Y^*(\textit{left}; T). \tag{6.42}$$

Summing the analogs of 6.41 for each term in 6.42 we get

$$\left| \sum_{j=1}^{4} \sum_{x_i \in \mathscr{X}_j} g(x_i) - 4 \frac{T}{24\sqrt{1+\alpha^2}} \text{area}(S) \right| \leq$$

$$\leq \sum_{j=1}^{4} \left| \sum_{x_i \in \mathscr{X}_j} g(x_i) - 4 \frac{T}{24\sqrt{1+\alpha^2}} \text{area}(S) \right| = O(T^{\gamma_0} \log T). \tag{6.43}$$

Note that

$$\text{meas}\{0 \leq t \leq T :\ L_\alpha(t) \in S\} = \left( \sum_{j=1}^{4} \sum_{x_i \in \mathscr{X}_j} f(x_i) \right) + O(1) =$$

$$= \sqrt{1+\alpha^2} \left( \sum_{j=1}^{4} \sum_{x_i \in \mathscr{X}_j} g(x_i) \right) + O(1). \tag{6.44}$$

Finally combining 6.43 and 6.44 yields

$$\left| \text{meas}\{0 \leq t \leq T :\ L_\alpha(t) \in S\} - \frac{T \text{area}(S)}{6} \right| = O(T^{\gamma_0} \log T), \tag{6.45}$$

which completes the proof of Case 1.

**Case 2:** $S$ is a subset of the front face such that $S \cap W = \varnothing$.

In Case 1 it was enough to take care of the contributions of the edge crossings with slope $1/\alpha$. To prove the analog for 6.45 for Case 2, we simply need to make the straightforward modification to include the contributions from edge crossings with slope $\alpha$ as well. As all convex test sets $S$ can be decomposed into test sets which satisfy either Case 1 or Case 2, together 6.45 and its analog for Case 2 prove Theorem 6. □

# CHAPTER 7

# GENERAL CUBE LINE DENSITY

From here on we no longer look at cube lines with any specific slope and look to prove uniform distribution for general cube lines. In this chapter we prove any cube line with an irrational slope will be dense over the cube. The argument used here is an adaptation of the argument by Katok and Zemlyakov [12].

Let $L(t) = L_\theta(t),\ t > 0$ be an infinite geodesic cube line with initial angle $\theta$ such that $L$ has an irrational slope. We prove $L(t)$ is dense by contradiction. If $L(t)$ is not dense, then the closure of the infinite trajectory,

$$\bar{L} = \text{ Closure}\{L(t) : t > o\}$$

is not the whole surface of the cube. There we can take a point $P_0 \in \partial\bar{L}$ on the boundary of $\bar{L}$—that is, a point such that every open neighborhood (with respect to the cube surface) of $P_0$ intersects both $\bar{L}$ and the complement of $\bar{L}$.

We will show that there is an open neighborhood of $P_0$ which is contained within $\bar{L}$, thus contradicting the fact that every open neighborhood about $P_0$ intersects the complement of $\bar{L}$. The closure $\bar{L}$ is a closed set which is invariant under the geodesic flow of initial angle $\theta$. Consider the geodesic, $\ell$, which starts at the point $P_0$. Since the slope of the cube line is irrational, there is at most one singular point along this line, thus the geodesic must be infinite in one of the two directions; without loss of generality, assume it is infinite in the forward direction. Let $I$ be an interval perpendicular to $\ell$ such that $P_0$ is the left endpoint of $I$, i.e.

$I = [P_0, R]$. We will show that there is a sub-interval $I^* = [P_0, Q] \subset I$ such that $I^* \subset \bar{L}$. By symmetry we can then do the same thing to the other side of $P_0$, which would yield an open neighborhood about $P_0$.

We claim that

$$\ell \text{ hits } I \text{ again.} \tag{7.1}$$

To prove 7.1, we first note that there is a sub-interval $I_0 = [P_0, R_0] \subset I$, such that no point in $I_0$ hits a singular point unless it has returned to hit a point in $I$ first. This is because the cube only has 8 vertices, and these half-lines from these singularities first intersect $I$ in at most $8 \times 4 = 32$ points. Thus if $I_0$ is short enough it will not contain any of these (at most) 32 points. We assume that $|I_0| < \frac{1}{2}|I|$ (where $|.|$ denotes interval length).

Now flow the interval $I_0$ forward with initial angle $\theta$, and we apply Poincare's recurrence theorem (proof first published by Caratheodory [18][4]), which works for any measure-preserving flow. This tells us that the flow of interval $I_0$ must return and overlap $I_0$. We can further consider the space of the surface of the cube as 48 faces, each one a surface of the cube with a particular orientation, or trajectory on the cube. This is still a measure-preserving flow and thus ensures that the interval $I_0$ must return and overlap $I_0$ when it is going in the same trajectory as $\ell$. Note in doing so we also must redefine $I_0$ so that no point hits a singular point unless it has first returned to hit a point in $I$ with the same trajectory as $\ell$.

Since $I_0 \subset I$, it follows that $I_0$ must return to $I$; consider the first time $I_0$ returns to $I$. It follows from the definition of $I_0$ that this flow must be a parallel distance-preserving flow. If $\ell$, starting from $P_0$, returns to $I$, then 7.1 is proved. Otherwise it follows that the other endpoint of $I_0$, $R_0$ returns to $I$ at a point $R_0' \in I$. We distinguish two cases.

**Case 1:** $R_0'$ is closer to endpoint $P_0$ than to $R_0$.

Since the trajectory starting at $R_0$ hits on one side of $P_0$ (the side within $I$), and the trajectory $\ell$ starting at $P_0$ hits outside of $I$ (which therefore is on the other side of $P_0$), there must be a point $R_1 \in I_0$ such that the trajectory starting at $R_1$ hits $P_0$. Now we consider the sub-interval $I_1 = [P_0, R_1] \subset I_0$, and repeat the previous argument for $I_1$. The geodesic flow from $I_1$ must return to $I_1$, and thus to $I$, and we know the first time this happens will be when $R_1$ hits $P_0$.

However Poincare's recurrence theorem says there will be not just one overlap but infinitely many, and since up to the first overlap none of $I_1$ hit $I \setminus I_0$, it follows that up to the second hit the flow will still be a parallel distance-preserving flow. If trajectory $\ell$ starting at $P_0$ returns to $I$ on the second overlap, we are done; otherwise the trajectory starting at $R_1$ hits $I$ on the second overlap. But since the trajectory from $R_1$ hit $P_0$ on the first overlap, this implies $\ell$ hits $I$, proving 7.1 in Case 1.

**Case 2:** $R_0'$ is closer to endpoint $R$.

In this case we replace $I$ with its one-third counterpart $I_{1/3} = [P_0, R_3] \subset I$ where $|I_{1/3}| = \frac{1}{3}|I|$. Again there is a sub-interval $I_{1/3}^{(0)} = [P_0, R_4] \subset I_{1/3}$ such that no point of $I_{1/3}^{(0)}$ hits a singularity without having first hit a point in $I_{1/3}$ with the same trajectory as $\ell$ (we again need only avoid a finite number of points, as above). Again we add the condition $|I_{1/3}^{(0)}| < \frac{1}{2}|I_{1/3}|$, with $|.|$ denoting length. The flow of $I_{1/3}^{(0)}$, until it returns to $I_{1/3}$, is a parallel distance-preserving flow. If $\ell$ starting from $P_0$ hits $I_{1/3}$ we are done. Otherwise the other endpoint $R_4$ returns to $I_{1/3}$ at a point $R_4'$. We again have two cases. If $R_4'$ is closer to endpoint $P_0$ of $I_{1/3}$, we are done by repeating the proof of 7.1 in Case 1 as above.

If $R_4'$ is closer to endpoint $R_3$, then since $|I_{1/3}^{(0)}| < \frac{1}{2}|I_{1/3}|$, $P_0$ must be outside $I_{1/3}$ to the right of $R_3$. But where $P_0$ hits is $\frac{1}{3}|I|$ right of $R_4'$, which is in $I_{1/3}$, and thus at most $\frac{1}{3}|I|$ right of $P_0$. Thus where $P_0$ hits is between $\frac{1}{3}|I|$ and $\frac{2}{3}|I|$ right of $P_0$, and is thus in $I$, proving 7.1.

This completes the proof of 7.1: $\ell$ hits $I$ again at some point $P_1 \neq P_0$.

If $[P_0, P_1] \in \bar{L}$ then we get our contradiction and we are done. Suppose it is not true. Then there is some point $P_2 \in [P_0, P_1]$ with $P_2 \notin \bar{L}$. Since $\bar{L}$ is closed there exists a largest open interval $J \subset [P_0, P_1]$ containing $P_2$ which is also disjoint from $\bar{L}$. Let $P_3$ be an endpoint of $J$, then $P_3 \in \bar{L}$. As the slope of the cube line is irrational, there is a direction in which we get an infinite geodesic $\ell'$ starting at $P_3$. From here we repeat the argument above for 7.1, simply replacing $(\ell, I)$ with $(\ell', J)$. Thus $\ell'$ will hit $J$ again, but this is a contradiction since $\bar{L}$ is flow-invariant with $J$ and $\bar{L}$ disjoint.

□

We note that the density proof extends to a much larger set of geodesics, far beyond those for which we can prove uniformity. For example, this density argument works on a rectangular

prism with a square base and height an irrational multiple of the base side length.

# CHAPTER 8

# CUBE SYMMETRIC UNIFORM DISTRIBUTION

In this chapter we will look at uniformity over general cube lines, but with respect to a particular class of test sets, cube symmetric sets, which will be described to follow.

We look at some appropriate projections of a cube line, $L(t)$ with starting location $L(0) = s = (1, s_1, s_2)$ on the front face and initial direction given by unit vector $(v_1, v_2)$, as $t$ runs in the interval $0 \leq t < \infty$, $t \to \infty$. We begin by examining the projection of $L(t)$ on to the $y$-axis. For small $t > 0$, its location is given by $s_1 + tv_1$, representing a motion with constant speed $v_1$. After $L(t)$ hits an edge, however, we have several different cases. By symmetry we can assume that $v_1, v_2 > 0$, thus $L(t)$ will first hit either the top or right edge, leaving us with two cases.

**Case 1:** $L(t)$ goes from the front face to the right face.

The switch to the right face happens along the $y$-axis, i.e. when $s_2 + tv_1 = 1$, or when $t = (1 - s_2)/v_1$. For $t > (1 - s_2)/v_1$, with $t - \frac{1-s_2}{v_1}$ small, then we consider the projection of $L(t)$ on the $x$-axis. It will move backwards along the unit interval $[0, 1]$ starting from 1 (as $x = 1$ is the front face) with the same constant speed $v_1$ for the duration that $L(t)$ is on the right face.

**Case 2:** $L(t)$ goes from the front face to the top face.

In this case the projection of $L(t)$ moves forward with the same constant speed $v_1$ as before, and the point where the equality $s_2 + tv_1$ occurs has several possibilities: it can occur on (1) the top face, (2) the back face, (3) the bottom face or (4) the front face. We consider what happens for each of these cases as the cube line travels along the next face—that is, again, $t > (1-s_2)/v_1$, with $t - \frac{1-s_2}{v_1}$ small, and consider the projection:

1. on the $z$-axis: The top face being $z = 1$, leaving the top face moves backwards along the unit interval $[0,1]$ starting from 1 with the same constant speed $v_1$ for at least the duration that $L(t)$ stays on the right face.

2. on the $x$-axis: The back face being $x = 0$, leaving the back face moves forwards along the unit interval $[0,1]$ starting from 0 with the same constant speed $v_1$ for at least the duration that $L(t)$ stays on the right face.

3. on the $z$-axis: The bottom face being $z = 0$, leaving the top face moves forwards along the unit interval $[0,1]$ starting from 0 with the same constant speed $v_1$ for at least the duration that $L(t)$ stays on the right face.

4. on the $x$-axis: The top face being $x = 1$, leaving the top face moves backwards along the unit interval $[0,1]$ starting from 1 with the same constant speed $v_1$ for at least the duration that $L(t)$ stays on the right face.

In order to simultaneously work with these projections onto different axes, and in different directions, going forward and backwards on the unit interval $0 \leq x \leq 1$ starting, respectively, at 0 or 1, we unify these cases by working on a "generic" unit interval $[0,1]$, where a billiard point on the interval represents the projection of the line onto whichever axis the line is traveling with a speed of $v_1$, and use the identity $x = 1 - x$. This identity leads to both backwards and forwards motion in the unit interval represented by first a backwards, and then a forwards motion of a point billiard in the half-interval $[0,1/2]$. Thus we have a billiard point moving at constant speed $v_1$ on the interval $[0,1/2]$, which bounces back at the endpoints 0 and $1/2$.

We next study the motion of $L(t)$ from a different viewpoint. We begin by looking at the projection of $L(t)$ on the $z$-axis at the start on the front face. When $t > 0$ is small, the projection

is $s_3 + tv_2$, representing a motion with constant speed $v_2$. Once $L(t)$ leaves the front face, we again have two cases. In the first case, the line $L(t)$ leaves the front face and goes to the top face. In this case the transition occurs at $s_3 + tv_2 = 1$, or $t = (1 - s_3)/v_2$, and so we change our projection onto the $x$-axis. The front being $x = 1$, for $t > (1 - s_3)/v_2$ with $t - (1 - s3)/v_2$ small, our projection moves backwards along the $[0,1]$ interval starting at 1, with constant speed $v_2$. In the second case, when $L(t)$ goes from the front face to the right face, we again have four possibilities for where $s_3 + tv_2 = 1$ can occur. Repeating the argument above, we can again combine all the cases into a single unified case where we work with another "generic" interval $0 \leq x \leq 1$, where a billiard point on the interval represents the projection of the line onto whichever axis the line is traveling with a speed of $v_2$, and again having the identity $x = 1 - x$ to unify the backwards and forwards movements into a single movement on the half-unit interval $[0, 1/2]$. Thus, again, we have a billiard point moving at constant speed, $v_2$, on the interval $[0, 1/2]$, which bounces back at the endpoints 0 and $1/2$.

Combining these two different one-dimensional views of the motion $L(t)$ of a particle on a geodesic of the cube surface, we obtain a single 2-dimensional representation of $L(t)$ defining by the mapping:

$$\Phi : L(t) \rightarrow \text{ constant speed point billiards motion on a square table.}$$

As both 1-dimensional movements are on the half-unit interval $[0, 1/2]$, the two dimensional movement is on the square table $[0, 1/2]^2$. Due to the identity $x = 1 - x$, the starting point of the billiard on this square table is:

$$(\min\{s_2, 1 - s_2\}, \min\{s_3, 1 - s_3\}),$$

and the initial velocity will be on of the four vectors $(\pm v_1, \pm v_2)$—all of which have constant unit speed.

By using a simple scaling argument to turn working with our $[0, 1/2]^2$ square into working with the unit square $[0,1]^2$ instead, we have reduced this problem to the familiar problem of billiard paths on the unit square discussed in Chapter 1, see Figure 1.3.

As $\Phi$ is a billiard movement with constant speed over a square table, the previously mentioned König-Szücs-Kronecker-Weyl theorem implies that if the slope of the movement, $v_1/v_2$, is rational it will be periodic, and if it is irrational it will be uniformly distributed over the square, $[0,1/2]^2$.

What does this result over the square $[0,1/2]^2$ tell us about the behavior of the geodesic over the surface of the cube? The primary issue is that the unifying process that allows us to project onto a single half-unit interval in a uniform way also resulted in the loss of information about where the geodesic is at any given time. Namely, we lost whether the projection onto $[0,1/2]$ is the projection onto the $x$, $y$, or $z$-axis, as well as the direction of the movement on the axis—whether it goes forward from 0 to 1, or backwards from 1 to 0. Despite this, we actually have maintained some information, and by reversing these unifying steps we can, for subsets of the square billiard table, $S \subset [0,1/2]^2$, determine precisely which subset of the cube's surface maps to the set $S$.

Let $\varepsilon > 0$ be arbitrarily small, but fixed, and consider the following two squares in the billiard table $[0,1/2]^2$:

$$S = [a-\varepsilon, a+\varepsilon] \times [b-\varepsilon, b+\varepsilon] \text{ and } S^* = [b-\varepsilon, b+\varepsilon] \times [a-\varepsilon, a+\varepsilon],\ \varepsilon \le a,b \le \frac{1}{2} - \varepsilon.$$

Consider $\Phi^{-1}(S \cup S^*)$; we claim that if $S$ and $S^*$ are not too close to each other–precisely that $|a-b| \ge \varepsilon$–then $\Phi^{-1}(S \cup S^*)$ consists of $6 \cdot 8 = 48$ disjoint congruent copies of the square $S$ on the surface of the unit cube.

Taking for example the front face of the cube, and examining what subsets of the front face $\Phi$ would map to $(S \cup S^*)$, we see that-

$$\text{front face} \cap \Phi^{-1}(S \cup S^*) =$$

$$= \{(1,y,z) : a-\varepsilon \le y \le a+\varepsilon, b-\varepsilon \le z \le b+\varepsilon\} \cup$$

$$\cup \{(1,y,z) : 1-a-\varepsilon \le y \le 1-a+\varepsilon, b-\varepsilon \le z \le b+\varepsilon\} \cup$$

$$\cup \{(1,y,z) : a-\varepsilon \le y \le a+\varepsilon, 1-b-\varepsilon \le z \le 1-b+\varepsilon\} \cup$$

$$\cup\{(1,y,z) : 1-a-\varepsilon \le y \le 1-a+\varepsilon, 1-b-\varepsilon \le z \le 1-b+\varepsilon\}\cup$$

$$\cup\{(1,y,z) : b-\varepsilon \le y \le b+\varepsilon, a-\varepsilon \le z \le a+\varepsilon\}\cup$$

$$\cup\{(1,y,z) : 1-b-\varepsilon \le y \le 1-b+\varepsilon, a-\varepsilon \le z \le a+\varepsilon\}\cup$$

$$\cup\{(1,y,z) : b-\varepsilon \le y \le b+\varepsilon, 1-a-\varepsilon \le z \le 1-a+\varepsilon\}\cup$$

$$\cup\{(1,y,z) : 1-b-\varepsilon \le y \le 1-b+\varepsilon, 1-a-\varepsilon \le z \le 1-a+\varepsilon\}. \tag{8.1}$$

This corresponds to the projection onto $[a-\varepsilon, a+\varepsilon]$ and onto $[b-\varepsilon, b+\varepsilon]$ from either the $y-$ or the $z$-axis, and from either end of the axis, i.e. both from $[a-\varepsilon, a+\varepsilon]$ and from $[1-a-\varepsilon, 1-a+\varepsilon]$. There are 8 copies of $S$ in 8.1, and the analogous versions of 8.1 for the other 5 faces of the cube comprise the other 40 copies.

We see from the above that $\Phi(S \cup S^*)$ is comprised of 48 disjoint congruent copies of the square $S$ on the surface of the unit cube, and, on the other hand, it is a well known fact that the symmetry group of the cube has 48 elements (24 rotational elements isomorphic to the 4! permutations of 4 different symbols, and their reflections). This is not a coincidence: the 48 copies of the square $S$ arise from applying each of the 48 elements of the symmetry group to a single copy,

$$\{(1,y,z) : a-\varepsilon \le y \le a+\varepsilon, b-\varepsilon \le z \le b+\varepsilon\}$$

of $S$ on the front face (or equivalently of any other copy of $S$). This means we may refer to the set $\Phi^{-1}(S \cup S^*)$ as an *elementary cube-symmetric set*—that is, $\Phi^{-1}(S \cup S^*)$ is invariant under any element of the isometry group of the cube. By using elementary cube-symmetric sets with arbitrarily small $\varepsilon$, we can well approximate every Jordan measurable cube-symmetric set by a finite union of elementary cube-symmetric sets. (We recall Jordan measurable sets are sets whose characteristic function is Riemann integrable, or, equivalently, those whose boundary has Lebesgue measure zero.)

If the slope of the geodesic, $v_1/v_2$, is irrational, then the billiard path on the square table $[0,1/2]^2$ is uniformly distributed. This means that the relative time the billiard spends in the

test set $S \cup S^*$ for $0 \leq t \leq N$, tends to (given that $[0,1/2]^2$ has area 1/4),

$$\frac{\text{area}(S \cup S^*)}{1/4} = \frac{2 \cdot 2\varepsilon \cdot 2\varepsilon}{1/4} = 32\varepsilon^2,$$

as the length $N$ of the time/arc-length interval tends to infinity. Using the inverse mapping $\Phi^{-1}$, it follows that the relative time the geodesic $L(t)$ spends in the symmetric test set $\Phi^{-1}(S \cup S^*)$ as $0 \leq t \leq N$ tends to the same value of $32\varepsilon^2$ as $N \to \infty$, which is the exact limit density of the uniform distribution on the cube,

$$\frac{\text{area}(\Phi^{-1}(S \cup S^*))}{6} = \frac{48 \cdot 4\varepsilon^2}{6} = 32\varepsilon^2,$$

or in other words, is the relative area of the cube symmetric test set to the cube surface.

As we can approximate every Jordan measurable cube-symmetric set by a finite union of elementary cube-symmetric sets, we get the following result.

**Theorem 7** (Weyl Type Uniformity Result)**.** *If $v_2/v_1$ is irrational, for every non-pathological starting point $L(0)$, the relative time the geodesic $L(t)$ spends in a given cube-symmetric Jordan measurable test set as $0 \leq t \leq N$, $N \to \infty$ tends to the limit density in the case of uniform distribution on the cube surface, i.e., tends to the relative area of the test set.*

□

A simple example of a cube-symmetric set for which this theorem applies is the $r$-neighborhoods of the centers of the six faces of the cube. For $0 \leq r \leq 1/2$, consider the set comprised of the union of the six circular discs of radius $r$ centered at each of the center points of the six faces of the cube.

Another example is the $r$-neighborhood of the edges of the cube. Again, for $0 \leq r \leq 1/2$, then the set comprised of all points that are within $r$ distance of any edge of the cube is a cube-symmetric set.

There are, of course, infinitely many examples, again comprising all sets which are invariant under all elements of the isometry group of the cube.

We note that Theorem 7 applies the Kronecker-Weyl theorem [20], and yields a Weyl type uniformity result, meaning it applies to Jordan measurable functions. Alternatively to this

we can combine the mapping $\Phi$ and the unfolding technique of König-Szücs with Birkhoff's ergodic theorem to get a Birkhoff type uniformity result, i.e. one which applies to Lebesgue measurable sets.

Applying Birkhoff's ergodic theorem (see 4.3) to the unfolded torus line of the $\Phi$ mapping we get the result.

**Theorem 8** (Birkhoff Type Uniformity Result)**.** *If $v_2/v_1$ is irrational, for almost every starting point $L(0)$, the relative time the geodesic $L(t)$ spends in a given cube-symmetric Lebesgue measurable test set as $0 \leq t \leq N$, $N \to \infty$ tends to the limit density in the case of uniform distribution on the cube surface, i.e., tends to the relative 2-dimensional Lebesgue measure of the test set.*

□

We present an alternative way to see the results in Theorem 7 and Theorem 8 without using projections to achieve the mapping $\Phi$. Let $L(t)$ be a cube line with initial direction $(v_1, v_2)$, and by symmetry assume $v_1, v_2 > 0$, and $L(0) = (1, s_2, s_3)$ is on the front face. We take the front face of the cube and generate 48 copies of it, corresponding to each of the 48 elements in the isometry group of the cube. When the cube line leaves the front face, it enters another face of the cube, but by applying one of the cube isometries, we can treat the cube line entering the new face as though it were entering the front face. We do this in the following way:

1. If the cube line is entering a new face because $s_2 + v_1 t = 0$ modulo 1, then apply the isometry that would make the cube line enter the front face from the left edge, with direction $(v_1, v_2)$.

2. If the cube line is entering a new face because $s_3 + v_2 t = 0$ modulo 1, then apply the isometry that would make the cube line enter the front face from the bottom edge, with direction $(v_1, v_2)$.

(8.2)

Note also that each isometry selected is unique. Rather than treating the cube line as though it re-enters the same face, we treat it as entering one of the 48 copies of the front face, whichever

corresponds the isometry applied to the cube as described in 8.2.

We repeat the above process whenever the cube line enters a new face; we select the isometry that (when applied to the original orientation of the cube) would follow the procedure in 8.2. In doing this we can see that the union of the lines drawn on each of the 48 copies of the front face together form a torus line on the unit square, $[0,1]^2$, with starting point $(s_2, s_3)$, and direction $(v_1, v_2)$. Or rather, another way to look at this is that lines drawn on the 48 copies of the front face partition a torus line on the unit square. Moreover, it is clear that any set on this square $[0,1]^2$ corresponds to a cube-symmetric set on the surface of the cube $[0,1]^3$, as the square $[0,1]^2$ was formed by taking a single face, and unioning it with copies of itself under all isometries of the cube.

From here the results follow as before, using either the Kronecker-Weyl theorem or the König-Szücs theorem on any test set for the torus line on the square $[0,1]^2$, we get Weyl and Birkhoff type uniformity, respectively, and using the fact that each set of the square $[0,1]^2$ corresponds to a cube-symmetric set of the surface of the unit cube $[0,1]^3$, we get the results of Theorem 7 and Theorem 8, respectively.

# CHAPTER 9

# GENERAL CUBE LINE UNIFORM DISTRIBUTION

To expand our results regarding cube-symmetric sets to results on all test sets, we use the discretization technique described in Chapter 2. Let $L(t)$ be a given cube line with irrational slope $\alpha$, and let $T = T_\alpha$ be the flow on interval of edge directions $S = [0, 24)$ induced by this cube line, as described in Chapter 2. Our goal is to show that if there exists a subset $S^* \subset S$ such that $0 < \mu(S^*) < 24$ and $S^*$ is invariant under $T$, then we have a contradiction. That is to say, our goal is to show $T$ must be ergodic, which would imply Weyl and Birkhoff type uniform distribution of the cube line on the cube's surface, as implied by Theorem 1 and Theorem 2 respectively.

We first recall a well known fact about continued fractions: the convergents $p_k/q_k$ of the irrational shift $\alpha$ represent a particularly good rational approximation of $\alpha$, which is to say,

$$\left|\alpha - \frac{p_k}{q_k}\right| \approx \frac{1}{q_k q_{k+1}}.$$

More precisely, we have the bounds

$$\frac{1}{q_k(q_{k+1}+q_k)} \leq \left|\alpha - \frac{p_k}{q_k}\right| \leq \frac{1}{q_k q_{k+1}}. \tag{9.1}$$

The power of the transformation $T$ are very closely related to the terms in the sequence of the $\alpha$ rotation, $m\alpha$ modulo 1 for integers $m$, and by 9.1 we get the approximation

$$\left| m\alpha - \frac{mp_k}{q_k} \right| \leq \frac{m}{q_k q_{k+1}} \leq \frac{\varepsilon}{q_k} \text{ as long as } 1 \leq m \leq \varepsilon q_{k+1}, \tag{9.2}$$

where $\varepsilon > 0$ is arbitrarily small but fixed.

The approximation in equation 9.2 motivates our decomposition of our interval of edge directions $S = [0, 24)$ into many very short sub-intervals of length $1/q_k$ (where we will eventually take the limit as $k \to \infty$ at the end of the proof). We introduce to this an unconventional, but very convenient terminology; for any integer $r \geq 1$, we call an interval of the form $[b/r, (b+1)/r)$, for some integer $b$, an *r-interval*, called this because it has length $1/r$.

Consider the decomposition of $S = [0, 24)$ into $24q_k$ different $q_k$-intervals, where $q_k$, as above, is a denominator of a convergent of $\alpha$. From the above equations, 9.1 and 9.2, it follows that, except for a $0(\varepsilon)$ error of terms, $T^n$ will map a $q_k$-interval of this decomposition to another $q_k$ interval of this decomposition, assuming $1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\}$. Note the implicit constant in the $O(.)$ term is absolute.

To break down this claim, let us look at an example of a $q_k$-interval. Let $I = [b/q_k, (b+1)/q_k)$ be a $q_k$ interval and, for simplicity assume $0 \leq b \leq q_k$—that is, the interval $I$ is in $[0, 1)$. Consider the sequence of images of $I$ under different powers of $T$, $TI, T^2I, T^3I, \ldots$. From the formulation of $T$, the image $TI$ will be an interval of length $q_k$, unless an endpoint of $TI$ is very close to singular point of $T$, i.e. where $T$ jumps between piecewise linear sections. This can occur in one of two ways, so we can formally write this scenario as (using $\{.\}$ to represent the fractional part):

$$\left\{ \frac{b}{q_k} + \alpha \right\} = 1 - \alpha + \text{ very small positive} \tag{9.3}$$

or

$$\left\{ \frac{b+1}{q_k} + \alpha \right\} = \text{ very small positive.} \tag{9.4}$$

In these cases, $TI$ is split into two intervals, and 9.1 tells us is that one of these intervals will be much shorter than the other (as the distance to the jump can be no more than the distance from $\alpha$ to $p_k/q_k$, and $p_k/q_k$ is, relatively, a very good approximation for $\alpha$). We call the much the

shorter part the *negligible part* and the longer part the *dominant sub-interval*. Intuitively, what we will do is discard the negligible part as, just as the name implies, negligible, and only use the dominating sub-interval. We will refer to the occurrence of 9.3 or 9.4 as a *truncation*.

Iterating this process through the powers of $T$, we see that if a truncation occurs at $T^m I$, $m \geq 1$, then we have analogous conditions for 9.3 and 9.4 as follows:

$$\left\{ \frac{b}{q_k} + m\alpha \right\} = 1 - \alpha + \text{ very small positive} \tag{9.5}$$

or

$$\left\{ \frac{b+1}{q_k} + m\alpha \right\} = \text{ very small positive.} \tag{9.6}$$

Similarly to above, when the truncation occurs, 9.2 tells us that once again one part will be much shorter than the other part; we will have a negligible part and a dominating sub-interval. More specifically, 9.2 implies that we can bound the size of the negligible as

$$\leq \frac{\varepsilon}{q_k} \text{ assuming that } m \leq \varepsilon q_{k+1}. \tag{9.7}$$

If $n \leq \min\{\varepsilon q_{k+1}, q_k\}$, then the sequence $TI, T^2 I, \ldots, T^n I$ can have at most 2 truncations, and thus have at most 2 negligible parts removed. Thus there is a loss of at most a $2\varepsilon$-part. Thus it follows that $T^n$ will "almost" map the $q_k$-interval $I$, and similarly all $q_k$-intervals, to another $1/q_k$ sized interval ("almost" here meaning with an relative error term of $O(\varepsilon)$). Finally, we note that as all $24q_k$ $q_k$-intervals are almost mapped to a $1/q_k$ sized interval of $[0,24)$, each must in fact be "almost" mapped to another $q_k$-interval, again with the same $O(\varepsilon)$ error term.

The next step is to apply Lebesgue's density theorem [16]. Let $\delta > 0$ be arbitrarily small but fixed, with relation to $\varepsilon$ specified later. Lebesgue's density theorem tells us that there exists a finite threshold value $k_0 = k_0(S^*; \delta)$ so that, for every $k \geq k_0$, the overwhelming majority of $q_k$ intervals $I$ in $[0,24)$ have the property that the $S$-density $q_k \cdot meas(I \cap S)$ is either greater than $1-\delta$, or less than $\delta$. We call the violator $q_k$-intervals $\delta$-violators. Precisely, Lebesgue's density theorem says the total number of $\delta$-violator $q_k$ intervals is less than $24q_k\delta$.

Let $0 < \eta < 1$ be another "small parameter" which we will specify later. A special case of

the $\delta$-violators is when they are $\eta$-violators, that is, for a $\delta$-violator $q_k$-interval, $I$,

$$\eta \leq q_k \cdot meas(I \cup S) \leq 1 - \eta.$$

Assume an $\eta$-violator exists, and let $I^*$ be one such $\eta$-violator $q_k$-interval—that is, it has $S$ density,

$$\eta \leq q_k \cdot meas(I^* \cup S) \leq 1 - \eta.$$

The general idea of what we will do next is show that there are a limited number of $\eta$-violators, and if an interval is an $\eta$-violator, then the interval it "almost" maps to must also be an $\eta$-violator. This will imply that that a path starting on an $\eta$-violator must hit these limited number of $q_k$-intervals many times. Hitting few $q_k$-intervals many times is a very unnatural path for the cube line, and as we will show is, in fact, impossible. This will provide a contradiction that will allow us to rule out the existence of $\eta$-violators.

The more precise argument is as follows:

We recall that for $n \leq \min\{\varepsilon q_{k+1}, q_k\}$, $T^n I^*$ is a well approximated by another $q_k$-interval, say $I_n^{**}$ (with an error term of $O(\varepsilon)$). Then we have that

$$\delta \leq \eta - O(\varepsilon) \leq q_k \cdot meas(I_n^{**} \cup S) \leq 1 - \eta + O(\varepsilon) \leq 1 - \delta,$$

so long as $\eta$ is sufficiently large with respect to $\varepsilon$ (larger than a large absolute constant multiple), and $\delta > 0$ is sufficiently small.

It thus follows from the bound on the number of $\delta$-violators that the total number of different $q_k$-intervals $I_n^{**}$, $1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\}$ is less than $2 \cdot 24 q_k \delta$.

For a value $x \in I^*$, we label $x$ *good* if

$$|\{1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\} : T^n x \in I_n^{**}\}| \geq \frac{1}{2} \min\{\varepsilon q_{k+1}, q_k\}.$$

Otherwise, we label $x$ as *bad*.

**Lemma 11.** *At least half of $x \in I^*$ are good.*

To prove Lemma 11, we use a standard "double counting average argument" as follows. Since $I_n^{**}$ is an $O(\varepsilon)$-approximation of $T_n I^*$ for each $1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\}$, we have

$$q_k \cdot meas(x \in I^* \ : \ T^n x \notin I_n^{**}) = O(\varepsilon),$$

and so we have, for some absolute constant $C$,

$$C \cdot \varepsilon \geq q_k \int_{x \in I^*} \frac{1}{\min\{\varepsilon q_{k+1}, q_k\}} |\{1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\} : \ T^n x \notin I^{**}\}| \geq$$

$$\geq q_k \int_{x \in I^* \text{ is bad}} \frac{1}{\min\{\varepsilon q_{k+1}, q_k\}} |\{1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\} : \ T^n x \notin I^{**}\}| \geq$$

$$\geq q_k \int_{x \in I^* \text{ is bad}} \frac{1}{2},$$

which moving the constants over yields,

$$meas(x \in I^* \text{ is bad}) \leq \frac{2C \cdot \varepsilon}{q_k}.$$

As $C$ is an absolute constant, by making $\varepsilon$ sufficiently small, we get

$$meas(x \in I^* \text{ is bad}) \leq \frac{1}{q_k},$$

completing the proof of the lemma. □

Since, for a given angle $\theta$, the set of pathological geodesics is countable, Lemma 11 implies we can find a *good* $x_0 \in I^*$ such that $T^n x_0, \ n \geq 0$ is on a non-pathological geodesic. As $x_0$ is *good* there are at least

$$\frac{1}{2} \min\{\varepsilon q_{k+1}, q_k\}$$

$n$'s in $1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\}$ such that $T^n x_0 \in I_n^{**}$. On the other hand we have bounded the number of different $q_k$-intervals which can be $I_n^{**}$,
$1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\}$, to be less than $2 \cdot 24 q_k \delta$. It thus follows from the pigeonhole principle

that $T^n x_0$ must visit some $q_k$-interval at least

$$\frac{\min\{\varepsilon q_{k+1}, q_k\}/2}{48 q_k \delta} \tag{9.8}$$

times as $n$ runs through $1 \leq n \leq \min\{\varepsilon q_{k+1}, q_k\}$.

As $\varepsilon$ has been fixed, we can make $\delta$ arbitrarily small. This, in turn, implies that $T^n x_0$ visits some particular $q_k$-interval far too many times. We next show that such a frequent visit is impossible. As previously shown in Chapter 8, for a cube-symmetric test set, the geodesic is uniform. This was done by reducing the case of the cube-symmetric test set to the problem of billiards on a square, which was further reduced to a torus line in a square. Uniformity of a torus line in a square is equivalent to uniformity of the irrational shift sequence, $n\alpha$, $n \geq 1$ modulo 1 for $\alpha$ irrational. To show that such frequent visits are impossible we look at the following lemma regarding the irrational shift sequence and compare it to the statement in 9.8. Note that this lemma retains the definitions of $q_k$ and $\alpha$ used throughout this chapter (i.e. $q_k$ is the denominator of a convergent of $\alpha$).

**Lemma 12.** *Let $I$ be an interval modulo 1 length $1/q_k$. The finite sequence $n\alpha$, $1 \leq n \leq N$ modulo 1 enters $I$ at most*

$$3\left(\frac{N}{q_k} + 1\right)$$

*times.*

If $N < q_k$ the statement of Lemma 12 is trivial, so assume $N \geq q_k$. Then we have $N/q_k \geq 1$, and we recall that

$$\left| n\alpha - \frac{np_k}{q_k} \right| \approx \frac{n}{q_k q_{k+1}},$$

which implies the set

$$\{n\alpha \text{ modulo } 1,\ 1 \leq n \leq N\}$$

is well approximated by the multiset consisting of the equidistant set

$$\{0, 1/q_k, 2/q_k, 3/q_k, \ldots, (q_k - 1)/q_k\}$$

with multiplicity either $\lfloor N/q_k \rfloor$ or $\lceil N/q_k \rceil$ (or simply $N/q_k$ when $N/q_k$ is an integer). As the

interval $I$ was of length $1/q_k$, it can contain at most 2 rational numbers with the same denominator $q_k$ and different integer numerators. As the multiplicity of each element in the above multiset is less than $(N/q_k)+1$, this completes the proof. □

To apply Lemma 12 to 9.8, we let $N = \min\{\varepsilon q_{k+1}, q_k\}$ and have $\varepsilon > 0$ be fixed. We then look at the cube symmetric version of our interval from 9.8 (i.e. the interval itself and all the rotations of it through the 24 symmetries that can be carried out in 3-space) and apply Lemma 12 to the torus line equivalent of this symmetric test set. We need an extra factor of 24 as the symmetric case includes 24 copies of the interval, but since $\delta > 0$ can still be set arbitrarily small this extra constant factor is not an issue, and thus Lemma 12 contradicts the existence of the test set described in 9.8.

We have thus shown that $\eta$-violator $q_k$-intervals *cannot* exist if $k$ is sufficiently large. This is summarized in the following lemma.

**Lemma 13.** *Let $\eta > 0$ be arbitrarily small but fixed. Then there exists a threshold $k_0 = k_(S;\eta)$ such that for $k > k_0(S;\eta)$ every $q_k$-interval $I$ in $[0,24)$ has the property that its S-density $q_k \cdot meas(I \cap S)$ is either $> 1-\eta$, or $< \eta$.*

□

For the rest of this chapter we make an assumption of $\alpha$, that the set of the continued fraction digits $a_k \geq 1$ of $\alpha$ is bounded (digit $a_k$ is also referred to as the $k^{th}$ partial quotient of $\alpha$). Formally, this means

$$\sup_{k\geq 1} \frac{q_{k+1}}{q_k} = A < \infty \text{ where } \alpha = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ldots}}}. \tag{9.9}$$

Combining the recurrence formula $q_{k+1} = a_k q_k + q_{k-1}$ with the bound in 9.9, we get the relationship

$$\sup_{k\geq 1} \frac{q_{k+1}}{q_k} \leq A+1. \tag{9.10}$$

We next need the following corollary of Lemma 13.

**Lemma 14.** *Let $\alpha$ be an irrational number with the properties described in 9.9, let*

$$0 < \eta \leq \frac{1}{2(A+2)},$$

*let $l > k > k_0(S;\eta)$ (see Lemma 13), let $I$ be a $q_k$-interval in $[0,24)$ with $S$-density $> 1-\eta$, let $J$ be a $q_l$-interval with $S$-density $< \eta$ with $J \subset I$. Then $J$ is in the $2\eta/q_k$-neighborhood of one of the endpoints of $I$.*

We prove Lemma 14 by contradiction; the basic idea of the proof is the bound on the continued fraction digits in 9.9 bounds the relative sizes of $q_i$-intervals and $q_{i+1}$-intervals, so given one $q_i$-interval and one $q_{i+1}$-interval intersect, with one having $S$-density $< \eta$ and the other $S$-density $> 1-\eta$, the intersection must be very small relative to the size of both intervals, and hence be near the edge of the intervals.

More precisely, suppose $(I;J)$ is a violator pair, and further assume $q_l$-interval $J$ has the minimum property that, for a fixed $q_k$-interval $I$, parameter $l$ is as small as possible. Let $J^*$ be a $q_{l-1}$-interval such that $|J^* \cap J| \geq |J|/2$. Then we have

$$\begin{aligned} meas(J^* \cap S) \leq meas(J \cap S) + meas((J^* \setminus J) \cap S) \leq \eta|J| + |J^* \setminus J| \leq \\ \leq \eta|J| + |J^*| - \frac{1}{2}|J| = \frac{\eta}{q_l} + \frac{1}{q_{l-1}} - \frac{1}{2}\frac{1}{q_l} \leq \frac{1-\eta}{q_{l-1}}, \end{aligned} \tag{9.11}$$

as by the bound in 9.10

$$\eta \leq \frac{1}{2(A+2)} \leq \frac{1/2}{\frac{q_l}{q_{l-1}}+1}.$$

Combining 9.11 and Lemma 13 we know the $S$-density of $J^*$ is less than $\eta$. The minimum property of $l$ implies $J^*$ cannot be a subset of $I$, as then $l-1$ would be a valid parameter, and thus $J^*$ must contain one of the endpoints of $I$.

On the other hand we also have

$$(1-\eta)|I| < meas(I \cap S) = meas(J^* \cap I \cap S) + meas((I \setminus J^*) \cap I \cap S) \leq$$

$$\leq meas(J^* \cap S) + |I \setminus J^*| < \eta|J^*| + |I| - |I \cap J^*|,$$

which implies that $|I \cap J^*| < \eta|I|$. When we combine this with the facts that: $J \subset I$, that $J$

intersects $J^*$, and that an endpoint of $I$ is contained in $J^*$; they together imply that $J$ must be contained within the $2\eta/q_k$ neighborhood of one of the endpoints, completing the proof of Lemma 14. □

For a given $q_k$-interval $I \subset [0,24)$, we define $I^\diamond$ to be the *middle* $(1-4\eta)$*-part sub-interval* of $I$. That is, $I^\diamond$ is the sub-interval of $I$ obtained by removing the interval of length $2\eta/q_k$ from each end of $I$.

We next have a quick lemma regarding these middle $(1-4\eta)$-part sub-intervals which is a corollary to the previous two.

**Lemma 15.** *Assume $\alpha$ is an irrational number with the properties described in 9.9, and let*

$$0 < \eta \leq \frac{1}{2(A+2)}.$$

1. *Let $k > k_0(S;\eta)$ (see Lemma 13), and let $I$ be a $q_k$-interval in $[0,24)$ with $S$-density $> 1-\eta$. Then $S$ contains the middle $(1-4\eta)$-part sub-interval $I^\diamond$ of $I$, except for a set of measure zero.*

2. *Similarly, let $I$ be a $q_k$-interval with $S$-density $< \eta$ and $k > k_0(S;\eta)$. Then $S$ is disjoint from the middle $(1-4\eta)$-part interval $I^\diamond$ of $I$, except for a set of measure zero.*

Combining Lemma 13 and Lemma 14 with $l \to \infty$ (i.e. using Lebesgue's density theorem[16]), we get that the $S$-density of the middle $(1-4\eta)$-part sub-interval $I^\diamond$ is arbitrarily close to 1 in the case of Lemma 15.1, and arbitrarily close to 0 in Lemma 15.2. Lemma 15 follows immediately from this. □

We note that the possible measure zero exception sets in Lemma 15.1 and Lemma 15.2 are irrelevant for our purposes; Theorem 1 and Theorem 2 both allow measure zero set exceptions, so for ease of notation we will assume $I^* \subset S$, and $I^\diamond \cap S = \varnothing$.

Now we define the two consecutive $q_k$-intervals

$$I_1 = \left[\frac{b}{q_k}, \frac{b+1}{q_k}\right) \text{ and } I_2 = \left[\frac{b+1}{q_k}, \frac{b_2}{q_k}\right)$$

in $[0,24)$, chosen such that both have $S$-density $> 1-\eta$. If 9.5 holds and $k > k_0(S;\eta)$ (see Lemma 13), then by Lemma 15 $S$ contains both middle $(1-4\eta)$-part
sub-intervals, $I_1^\diamond$ and $I_2^\diamond$, for these two sets. We define $I^*$ to be the longest sub-interval of $S$ that contains $I_1^\diamond$, and similarly define $I_2^*$ to be the longest sub-interval of $S$ which contains $I_2^\diamond$. Lastly, define $Right_1$ to be the right endpoint sub-interval of $I_1^*$, and $Left_2$ to be the left endpoint of the sub-interval $I_2^*$.

Suppose

$$Right_1 < Left_2. \tag{9.12}$$

We show that 9.12 is impossible if we choose $\eta$ to be sufficiently small in relation to $A$ (to be specified later, see 9.23).

The gap $Left_2 - Right_1$ is between two consecutive reciprocals of convergent denominators,

$$\frac{1}{q_{l-1}} > Left_2 - Right_1 \geq \frac{1}{q_l} \tag{9.13}$$

$Right_1$ must be contained in some $q_l$-interval $J_1$ such that $J_1$ has $S$-density $< \eta$, since $Right_1$ is the right endpoint of $I_1^*$. As $I_1^*$ is contained in $S$, this implies that $|I_1^* \cap J_1| < \eta |J_1|$.

By a similar argument, since $Left_2$ is the left endpoint of $I_2^*$, it must be contained in some $q_l$-interval $J^{(2)}$, which has $S$-density $< \eta$. Again, as $S$ contains $I_2^*$, we have that $|I_2^* \cap J^{(2)}| < \eta |J^{(2)}|$.

It may be that $J_1 = J^{(2)}$, or they may be different. If they are different, then there is a chain of consecutive $q_l$-intervals $J_1, J_2, \ldots, J_m = J^{(2)}$. Moreover, 9.10 and 9.13 imply that $m \leq A+1$. This implies that

$$\left| (Left_2 - Right_1) - \frac{m}{q_l} \right| < \frac{2\eta}{q_l} \text{ with some integer } 1 \leq m \leq A+1. \tag{9.14}$$

By repeating the argument above with $l+1$ rather than $l$, we get an analogous result to 9.14:

$$\left| (Left_2 - Right_1) - \frac{M}{q_{l+1}} \right| < \frac{2\eta}{q_{l+1}} \text{ with some integer } 1 \leq M \leq (A+1)^2. \tag{9.15}$$

By combining 9.14 and 9.15 into a single equation we get

$$\left|\frac{m}{q_l}-\frac{M}{q_{l+1}}\right|<\frac{2\eta}{q_l}+\frac{2\eta}{q_{l+1}} \text{ with some integers } 1\le m\le A+1, M\le (A+1)^2. \tag{9.16}$$

Multiplying 9.16 through by $q_{l+1}/m$ yields

$$\left|\frac{q_{l+1}}{q_l}-\frac{M}{m}\right|<\frac{4\eta q_{l+1}}{q_l m}<4\eta(A+1) \text{ with some integers } 1\le m\le A+1, M\le (A+1)^2. \tag{9.17}$$

Note, however, that

$$\frac{q_{l+1}}{q_l}=\frac{a_l q_l+q_{l-1}}{q_l}=a_l+\frac{1}{q_l/q_{l-1}}=a_l+\frac{1}{\frac{1}{a_{l-1}+\frac{1}{a_{l-2}+\cdots}}}.$$

This means that $q_{l+1}/q_l$ has bounded continued fraction digits, $a_{l-1}, a_{l-2}, \ldots,$
namely they are all less than $A$ (we note that $a_l$, the integer part of $q_{l+1}/q_l$, is also bounded by $A$, but is irrelevant). This brings us to a concept from Diophantine approximations which, vaguely summarized, says that if a real number (whether rational or irrational) has "small" continued fraction digits, then it cannot be well approximated by a small fraction unless it itself is a small fraction; in this case, $q_{l+1}/q_l$ is being well approximated by the small fraction $M/m$. This concept motivates our next step which is to examine a classical result in number theory.

**Diophantine Approximation Fact:** If a real number $x$ is approximated by a rational number $M/m$ such that:

$$\left|x-\frac{M}{m}\right|<\frac{1}{2m^2},$$

then $M/m$ is a convergent of $x$. Moreover, there is a convergent $P_i/A_i$ of $x$ such that $Q_i\le m<Q_{i+1}$, and $M/m=P_i/Q_i$.

By applying 9.17 with $x=q_{l+1}/q_l$, we see that

$$\left|\frac{q_{l+1}}{q_l}-\frac{M}{m}\right|=\left|x-\frac{M}{m}\right|<\frac{1}{2m^2}, \tag{9.18}$$

if given that we also have

$$4\eta(A+1) \leq \frac{1}{2(A+1)^2} \leq \frac{1}{2m^2}. \tag{9.19}$$

However, 9.19 can easily be guaranteed by setting

$$\eta \leq \frac{1}{8(A+1)^3} \tag{9.20}$$

Combining the Diophantine Approximation Face with 9.18 implies that $M/m$ is a convergent $P_i/Q_i$ of $x$ such that $Q_i \leq m \leq A+1$. This allows us to rewrite 9.17 as

$$\left| x - \frac{P_i}{Q_i} \right| < 4\eta(A+1) \text{ with some } 1 \leq Q_i \leq A+1. \tag{9.21}$$

On the other hand, 9.1 (which holds for both rational and irrational numbers) implies that

$$\frac{1}{Q_i(Q_{i+1}+Q_i)} \leq \left| x = \frac{P_i}{Q_i} \right|. \tag{9.22}$$

Applying our bounds $Q_i \leq A+1$ and $Q_{i+1} \leq (A+1)^2$ to the left hand side of 9.22, and then combining it with 9.21 yields the inequality

$$\frac{1}{(A+1)^2(A+2)} \leq 4\eta(A+1),$$

which yields a contradiction for sufficiently small $\eta$, namely letting

$$\eta \leq \frac{1}{4(A+1)^3(A+2)}. \tag{9.23}$$

We also note that this selection of $\eta$ will also satisfy the required bound in 9.20. This contradicts 9.9, and thus tells us that

$$Left_2 \leq Right_1. \tag{9.24}$$

A symmetric argument proves the analogous result if $I_1$ and $I_2$ both have $S$-density $< \eta$.

We next check the case where $I_1$ and $I_2$ are consecutive $q_k$-intervals in $[0,24)$—that is,

$$I_1 = \left[ \frac{b}{q_k}, \frac{b+1}{q_k} \right) \text{ and } I_2 = \left[ \frac{b+1}{q_k}, \frac{b+2}{q_k} \right),$$

Such that $I_1$ has $S$-density $> 1-\eta$ and $I_2$ has $S$-density $< \eta$. If 9.5 holds and $k > k_0(S;\eta)$ (see Lemma 13), then by Lemma 15 $S$ contains the middle $(1-4\eta)$-part sub-interval $I_1^{\diamond}$, and $S^c$, the complement of $S$, contains the middle $(1-4\eta)$-part sub-interval $I_2^{\diamond}$. Defining $I_1^*$ as before, the longest sub-interval of $S$ which contains $I_1^{\diamond}$, and defining $I_2^*$ to be the longest sub-interval of $S^c$ which contains $I_2^{\diamond}$, let $R_1$ denote the right endpoint of $I_1^*$ and $L_2$ denote the left endpoint of $I_2^*$.

From this we can repeat the argument from 9.12-9.23 to these definitions (bearing in mind that as $T$ is measure preserving, $S$ being invariant under $T$ implies $S^c$ is invariant under $T$ as well, and thus we can apply all the same arguments to it), and again we obtain that $R_1 < L_2$ is false, but as $I_1^*$ and $I_2^*$ must clearly be disjoint we get that

$$\text{the right endpoint of } I_1^* \text{ coincides with the left endpoint of } T_2^*. \tag{9.25}$$

We of course get analogous results in the symmetric case that $I_1$ has $S$-density $< \eta$ and $I_2$ has $S$-density $> 1-\eta$.

Combining the results of 9.24 and 9.25 we see that there can be no gaps between the sub-intervals fully contained in $S$ and the sub-intervals fully contained in $S^c$ found in each middle $(1-4\eta)$-part sub-intervals of the $q_k$-intervals. This tells us that any invariant set $S$ must have a very simple structure: it must be the finite union of sub-intervals. But this is impossible. If it were possible, the non-singular interval endpoints must be transformed to each other by $T = T_\alpha$. But $T$ is essentially an irrational (that has been partitioned onto the different directed edges), and so the iterated $T$-images of an endpoint modulo 1 must be the *infinite* set $x, x+\alpha, x+\alpha^2, \ldots$ modulo 1, but this contradicts the fact that $S$ is comprised of a *finite* union of intervals.

More formally, assume $S$ is the union of $W < \infty$ disjoint intervals, so there are a total of $2W$ interval endpoints. Let $x \in [0,24)$ be an interval endpoint such that $\{x\} \neq \{-m\alpha\}$, $m \geq 0$ an integer. Consider the sequence

$$\{x\}, \{x+\alpha\}, \{x+\alpha\}, \{x+3\alpha\}, \ldots, \{x+2W\alpha\}. \tag{9.26}$$

As $1-\alpha, 0, 1$ are not elements in the sequence 9.26, and the sequence is finite, there is a "buffer zone", i.e., there is some value $\zeta$, $0 < \zeta < \min\{\alpha, 1-\alpha\}$ such that the intervals $[0,\zeta], [1-\zeta, 1]$,

and $[1-\alpha-\zeta, 1-\alpha+\zeta]$ are disjoint from the sequence 9.26. We select a value $\zeta$ so that it satisfies these conditions, and so that $\zeta$ is smaller than the length of any of the $W$ intervals in $S$, and smaller than the length of any of the (at most $W+1$) intervals that form $S^c$. Consider the interval $I(x;\zeta) = [x-\zeta, x+\zeta]$, the interval of radius $\zeta$ about the point $x$. As $I(x;\zeta)$ has radius $\zeta$, we first know that none of the singular points $1-\alpha, 0, 1$ are in $I(x;\zeta)$, and similarly are not in any of the images of $I(x;\zeta)$, $T^j I(x;\zeta),\ 1 \leq j \leq 2W$.

Secondly, as $x$ is an endpoint of an interval in $S$, on one side of $x$ will be an interval in $S$, and the other side an interval in $S^c$. As $I(x;\zeta)$ extends on both sides of $x$ with a radius smaller than any interval in $S$ or $S^c$, this implies that half of $I(x;\zeta)$ will be in $S$ and the other half will be in $S^c$. We denote this property by referring to $I(x;\zeta)$ as a *half-and-half* interval. We further see that by the selection of $\zeta > 0$ being sufficiently small, all images of $I(x;\zeta)$, $T^j I(x;\zeta),\ 1 \leq j \leq 2W$, will also be *half-and-half* intervals. Thus the center points of $T^j I(x;\zeta),\ \leq j \leq 2W$ must all be endpoints of intervals comprising $S$. But they also have fractional part $\{x+j\alpha\}$, as described in 9.26, which must all be different since $\alpha$ is irrational. This implies all $2w+1$ center points (and thus interval endpoints) are unique, but this contradicts the fact that there are only $2W$ endpoints.

Next we handle the cases where $x \in [0,24)$ is an interval endpoint such that $\{x\} = \{-m\alpha\}$ for some integer $m \geq 2$. Then we can use the exact same argument as above by working with the inverse transformation $T^{-1}$, i.e. by using $T^{-j} I(x;\zeta)$, $0 \leq j \leq 2W$.

Thus we have that the only places that can be an endpoint of an interval in the decomposition $S$ are points $x \in [0,24)$ where $\{x\} = 0$ or $\{x\} = \{\alpha\} = 1-\alpha$, i.e. every endpoint in the interval decomposition is a singular point. Since $S$ is nontrivial, there is an integer $r$, $1 \leq r \leq 24$ such that we have either

$$(r-1, r-\alpha) \subset S \text{ and } (r-\alpha, r) \cap S = \varnothing, \tag{9.27}$$

or

$$(r-1, r-\alpha) \cap S = \varnothing \text{ and } (r-\alpha, r) \subset S. \tag{9.28}$$

Suppose that:

$$0 < \alpha < 1/2 \text{ and 9.27 holds.} \tag{9.29}$$

Suppose, without loss of generality, that $(r-1,r)$ represents the right edge of the front face, with the direction of the cube line traveling from the right face to the front face. Denote this directed edge $E = E(0,1)$. As $0 < \alpha < 1/2$, we know that $T = T_\alpha$ will map $E(0,1-\alpha)$ to the directed edge going from the front face to the right face; denote this edge $E' = E'(0,1)$. More precisely, $E(0,1-\alpha)$ will be mapped to $E'(\alpha,1)$. As $E(0,1-\alpha)$ is in $S$ (assumed by 9.29), $E'(\alpha,1)$ is in $S$ as well. Moreover, as $0 \leq \alpha \leq 1/2$, we know that $E'(\alpha,1)$ intersects both $E'(0,1-\alpha)$ and $E'(1-\alpha,1)$. As endpoints can only occur at singular points, this implies that the entire directed edge of $E'$ is in $S$. Repeating this argument (it can now be repeated on both $E'(0,1-\alpha)$ and $E'(1-\alpha,1)$), we quickly see that under condition 9.29

$$S \text{ must contain all 24 directed edges of the cube,} \tag{9.30}$$

which is impossible, given we assume $S$ is non-trivial.

The underlying reason this works, that is, the reason the argument above
spreads to all edges of the cubes, is the *connectivity* of a graph associated with the interval exchange transformation $T = T_\alpha : [0,24) \to [0,24)$. We define a directed graph $\vec{\mathbf{G}} = \vec{\mathbf{G}}_{cube}(\alpha)$ such that the vertices of $\vec{\mathbf{G}}$ are the 24 directed edges of the unit cube, and, for two vertices $E, E'$ in the vertex set of $\vec{\mathbf{G}}$, $(E,E')$ is an edge of $\vec{\mathbf{G}}$ iff there exists $x \in E$ such that $Tx \in E'$—that is, $T$ maps part of $E$ to $E'$. We note because of the structure of $T$, each vertex of $\vec{\mathbf{G}}$ has in degree 2 and out degree 2.

Ignoring the directions of the edges in $\vec{\mathbf{G}}$, we get a 4-regular (undirected) graph $\mathbf{G} = \mathbf{G}_{cube}(\alpha)$. We refer to $\vec{\mathbf{G}}$ as the *directed cube-surface-reachability graph*, and refer to the undirected version, $\mathbf{G}$, as the *cube-surface-reachability graph*.

We prove the following lemma to show that all 24 directed edges are reachable.

**Lemma 16.** *If $\alpha$ is irrational, then the cube-surface-reachability graph $\mathbf{G}$ is connected.*

As $\mathbf{G}$ is a relatively small concrete graph with 24 vertices and 48 edges, it can easily be verified with brute force checking. □

We note also a "clever" proof of Lemma 16 will be proved later (see Lemma 18).

We apply Lemma 16 as follows: under the conditions in 9.29 we have that $E(0,1-\alpha) \in S$,

and Lemma 16 implies

$$\vec{\varepsilon}(0, 1-\alpha) \subset S$$

is true for every directed edge $\vec{\varepsilon}$ in the cube. By applied Lemma 16 one more time, we can upgrade this to

$$\vec{\varepsilon} \subset S$$

for every directed edge $\vec{\varepsilon}$ in the cube, which yields 9.30.

We next verify the alternative choices to the assumptions made in 9.29 as symmetric cases.

If $1/2 < \alpha < 1$ and 9.27 holds, then the argument above holds perfectly, and yields an identical result to 9.30–$S$ must contain all directed edges.

If $0 < \alpha < 1/2$ and 9.28 holds, then the argument above holds for $S^c$, the complement of $S$. It shows $S^c$ must contain all directed edges of the cube, and thus $S$ is still trivial.

Finally, if $1/2 < \alpha < 1$ and 9.28 holds, the same argument as the previous case holds with $S^c$, and we get that again $S^c$ must contain all directed edges, and thus $S$ is trivial.

Since in all four cases we have obtained the contradiction that $S$ is trivial, this shows that there are no subsets of $S \subset [0, 24)$, with $0 < meas(S) < 24$, which are invariant under $T = T_\alpha$, for all irrational numbers $\alpha$ with bounded continued fraction digits.

We note this result alone is very interesting, as it includes many explicit famous numbers, namely the whole class of quadratic irrationals, including $\sqrt{2}$ and the previously mentioned $-2+\sqrt{5}$.

We next look at the other case, the set of irrationals $\alpha$ with unbounded continued fraction digits. The principal idea is the same, however: apply Lebesgue's density theorem[16], and extend it by utilizing "substantially overlapping intervals".

# CHAPTER 10

# UNIFORM DISTRIBUTION: UNBOUNDED DIGITS

Suppose $\alpha$ is an irrational number with unbounded continued fraction digits. Then there is a set of positive integers $K$ such that

$$\lim_{k\to\infty \text{ with } k\in K} \frac{q_{k+1}}{q_k} = \infty. \tag{10.1}$$

If $q_{k+1}/q_k$ is large, then 9.1 is an *extremely* good approximation of $\alpha$. Because of this, the argument for $\alpha$ with unbounded continued fraction digits is in fact quite a bit simpler than the argument for $\alpha$ with bounded digits.

We find it convenient here to rewrite the non-trivial, invariant under $T = T_\alpha$, set $S$ as follows. First we define a function,

$$F = F_s : [0,1) \to \{0,1\}^{24},$$

which we think of as a mapping from the 1-dimensional unit torus line $[0,1)$, to the set of $0-1$ sequences of length 24, and essentially turn $F$ into an indicator function for $S$ on each of the 24 directed edges. We define $F(x)$ by the following criteria; for integer $r$, $1 \le r \le 24$, the $r^{th}$ coordinate of $F(x)$ equals 1 if and only if $r-1+x \in S$.

We next use the tnterval exchange transformation $T = T_\alpha$ to define a second function

$$\mathbf{P}(x) = \mathbf{P}_\alpha(X) : [0,1) \to S_{24},$$

where $S_{24}$ is the group of permutations with 24 elements. We define $\mathbf{P}(x)$ so that: if $r$ and $s$ are two integers, $1 \leq r < s \leq 24$, and $T(r-1+x) \in [s-1, s)$, then the permutation $\mathbf{P}(x)$ sends $r$ to $s$. That is, $\mathbf{P}(x)$ is a permutation of the directed edges defined by, for a point $x$ on a directed edge, whichever directed edge $T$ sends $x$ to. We note that $\mathbf{P}(x)$ has only two possible values for a given value $\alpha$, as it is constant on $[0, 1-\alpha)$ and on $[1-\alpha, 1)$. We note importantly that $S$ being invariant under $T$ implies that $F$ satisfies the equation

$$F(x+\alpha) = \mathbf{P}(x)F(x) \text{ for every real } x. \tag{10.2}$$

From this it is clear that the set $S \subset [0, 24)$ is trivial if and only if the function $F_S$ is constant almost everywhere on $[0, 1)$ with value $\{0\}^{24}$ or $\{1\}^{24}$.

As previously, we begin by considering the partition with denominator $q_k$, or more precisely, consider the equidistant partition

$$\{0, 1/q_k, 2/q_k, 3/q_k, \ldots, (q_k-1)/q_k\} \tag{10.3}$$

of the unit torus $[0, 1)$. However, motivated by the $\alpha$-shift and its two special intervals $[0, 1-\alpha)$, $[1-\alpha, 1)$ with endpoints $0, 1-\alpha, 1$, we additionally examine another partition of the unit torus $[0, 1)$ as well. Let $\mathscr{P}_k(\alpha)$ denote the partition of the unit torus line $[0, 1)$, interpreted as a circle, with division points $\{-\ell\alpha\}$, $0 \leq \ell \leq q_k$, where $k \in K$, and once again $\{.\}$ represents the fractional part of a real number. We note that the values $\ell = 0, 1$ in $\{-\ell\alpha\}$ represent the endpoints of the special intervals $[0, 1-\alpha),\ [1-\alpha, 1)$.

For a fixed $\varepsilon > 0$, if $k \in K$ is sufficiently large, then $q_{k+1}/q_k$ is very large, and so the ratio of the gaps of the partition $\mathscr{P}_k(\alpha)$ will lie between $1-\varepsilon$ and $1+\varepsilon$, as (vaguely speaking) the number of partition points of $\mathscr{P}_k(\alpha)$ increases it (being an irrational shift) will resemble a uniform distribution over $[0, 1)$. More formally, the following corollary of 9.1

$$\left| \ell\alpha - \frac{\ell p_k}{q_k} \right| \leq \frac{\ell}{q_k q_{k+1}} < \frac{1}{q_{k+1}} \text{ as long as } 1 \leq \ell < q_k,$$

combined with 10.1, implies that the gaps of partition $\mathscr{P}_k(\alpha)$ have lengths in the range

$$(1-\varepsilon)\frac{1}{q_k} < \frac{1}{q_k} - \frac{1}{q_{k+1}} \leq \text{ length of a gap } \leq \frac{1}{q_k} + \frac{1}{q_{k+1}} \leq (1+\varepsilon)\frac{1}{q_k}.$$

Moreover, this means that, under the conditions in 10.1, the partition $\mathscr{P}_k(\alpha)$ becomes nearly identical to the partition in 10.3, with a relative error term of arbitrarily small $\varepsilon > 0$.

We note here that in the case of bounded continued fraction digits, we had to work with the partition in 10.3, because the similarities between it and $\mathscr{P}_k(\alpha)$ depend entire on condition 10.1—that is, that the continued fraction digits are unbounded. In this chapter however we can use the more convenient $\mathscr{P}_k(\alpha)$. One technical advantage of working with $\mathscr{P}_k(\alpha)$ is that an $\alpha$ (modulo 1) shift can be expressed on it without the need of any error term at all.

Let $I(\ell;2)$ denote the union of the two consecutive intervals (again interpreting the torus line $[0,1)$ as a circle) that share the endpoint $\{-\ell\alpha\}$, $0 \leq \ell < q_k$. This definition makes the overlap between two neighboring intervals $I(\ell;2)$ quite substantial—in fact, the overlapping portion will be nearly half the size of each interval (more precisely, the ratio between the size of the overlapping portion and the interval itself is at least $\frac{1-\varepsilon}{2}$). This makes our method of extending Lebesgue's density theorem by using overlapping intervals quite simple here.

We define the following terms:

$$\mathbf{P}^{(1)} = (x), \;\; \mathbf{P}^{(2)} = \mathbf{P}(x+\alpha)\mathbf{P}(x), \;\; \mathbf{P}^{(3)} = \mathbf{P}(x+2\alpha)\mathbf{P}(x+\alpha)\mathbf{P}(x),$$

and in general, for $r \geq 3$, we define

$$\mathbf{P}^{(r)}(x) = \mathbf{P}(x+(r-1)\alpha)\mathbf{P}^{(r-1)}(x).$$

Notice that $\mathbf{P}^{\ell-1)}(x)$ is constant on $I(\ell;2)$ for $2 \leq \ell < q_k$, because the only places $P(x)$ changes is at $0, 1-\alpha$, and 1, which are found at $\ell = 0, 1$.

From 10.2, we have that

$$F(x+(\ell-1)\alpha) = \mathbf{P}^{(\ell-1)}(x)F(x) \text{ for every real } x. \tag{10.4}$$

Note the translated copy $I(\ell;2)+(\ell-1)\alpha$ (modulo 1) of $I(\ell;2)$ is almost exactly the same as $I(1;2)$—or more precisely, it is an approximation with an error term of order $\varepsilon$. For $k \in K$ sufficiently large, by Lebesgue's density theorem [16] there exists $\ell_0$ in $1 \leq \ell_0 < q_k$ such that $F(x)$ is almost constant in the interval $I(\ell_0;2)$—or more precisely, that $F(x)$ is constant on a subset of $I(\ell_0;2)$ with measure $\geq (1-\varepsilon)|I(\ell_0;2)|$.

Moreover, combining 10.4 with the fact that $F(x)$ is almost constant in the interval $I(\ell_0;2)$ implies that $F(x)$ is almost constant in the interval $I(1;2)$—or more precisely, that $F(x)$ is constant on a subset of $I(\ell_0;2)$ with measure
$\geq (1-\varepsilon)^2|I(\ell_0;2)|$.

We next use 10.4 "backwards", by denoting $\mathbf{P}^{(r)*}(x)$ to be the inverse permutation of $\mathbf{P}^{(r)}(x)$. Then, by 10.4, we get

$$\mathbf{P}^{(\ell-1)*}(x)F(x+(\ell-1)\alpha) = F(x) \text{ for every real } x. \tag{10.5}$$

Using 10.5 we can translate this result to a general interval $I(\ell;2),\ 2 \leq \ell < q_k$, and obtain that $F(x)$ must be almost constant on each of them—or more precisely, $F(x)$ is constant on a subset of $I(\ell;2),\ 2 \leq \ell < q_k$ with measure $> (1-\varepsilon)^3|I(\ell;2)$. Since we know this for $\ell = 1$ as well, we now know $F(x)$ is almost constant on all intervals $I(\ell;2),\ 1 \leq < q_k$.

We now make use of the fact that the intervals $I(\ell;2)$ are comprised of consecutive intervals of $\mathscr{P}_k(\alpha)$. Thus we can take the $q_k$ different intervals $I(\ell;2)$ and line them up in such a way that between any two of them there is substantial overlapping (nearly 1/2 the interval), and together they will form a chain that covers the entire unit interval starting at 0 and ending at 1. As each interval $I(\ell;2)$ is constant on $(1-\varepsilon)^3$ portion of the interval, and it overlaps with its neighbors for nearly 1/2 $(-\varepsilon/2)$ its length, $F(x)$ must be almost constant with the same value on each interval and its on neighbors. As the intervals in the chain form the entire unit interval $[0,1)$, $F(x)$ must be constant with the same value for every interval. Thus $F(x)$ is constant on a subset of size $(1-\varepsilon)^3$.

Taking $\varepsilon \to 0$, we conclude that $F(x) = F_S(x)$ must be constant on $[0,1)$ almost everywhere. Constant in this case means almost every value of $x$ has the some $0-1$ sequence. This means the directed edges which have a 0 value in their coordinate of that sequence are almost never

visited, while the directed edges which have a 1 value in their coordinate are visited almost everywhere. But by Lemma 16, if one edge is visited almost everywhere, they all must be. Thus the constant value must be either $\{0\}^{24}$ or $\{1\}^{24}$. This contradicts the assumption that $S$ is non-trivial, and thus a non-trivial invariant under set $T$ cannot exist. We have therefore proved the following lemma.

**Lemma 17.** *Consider the interval exchange transformation $T = T_\alpha : [0,24) \to [0,24)$ defined in 2.3. If $\alpha$ is irrational then $T$ is ergodic.*

□

By Lemma 1 and Lemma 2, Lemma 17 completes the proof of Theorem 1 and Theorem 2. □

Lemma 17 represent the central technical part of Theorem 1 and Theorem 2. It is also somewhat similar to the crucial missing part in Gutkin's paper [8] (in the special case of cube surface), what he refers to as "an unpublished result of W. Veech" (see page 581 in [8]). As we mentioned at the beginning of Chapter 1, the proof of this crucial part in [8] remains unpublished since (email communication by W. Veech, Summer of 2016).

**Unique ergodicity.** We conclude this chapter with an upgrading of Theorem 1: we replace the "almost every starting point" exception to the Theorem with "every non-pathological starting point", just like we had in Lemma 1. It is based off the fact that the dynamical system $([0,24), T_\alpha)$ is *uniquely ergodic*. The concept of *uniquely ergodic* dynamical systems $(Y,T)$ (where $Y$ is compact and $T : Y \to Y$) was introduced by Furstenberg in 1972. Note here that we interpret $[0,24)$ as a circle (or rather the union of 24 separate circles), and thus it is compact. We cannot directly apply Furstenberg's general theorems to this problem, but instead we adapt his proof arguments. (See Sections 3.2 and 3.3 in his book [6]).

Let $\alpha$ be an irrational real number and let $\lambda(.)$ denote the one-dimensional Lebesgue measure.

**Step 1:** *The only $T_\alpha$-invariant Borel probability measure of the dynamical system $([0,24), T_\alpha)$*

*is* $\lambda(.)/24$ *(where of course* $T_\alpha = T_\alpha(cube)$*).*

This uniqueness property of a dynamical system is what Furstenberg refers to as *uniquely ergodic*. To prove Step 1, i.e. to prove $([0,24),T_\alpha)$ is *uniquely ergodic*, we rewrite $([0,24),T_\alpha)$ as a *skew product*. Let $R_\alpha : x \to x+\alpha$ (modulo 1) denote the irrational rotation by $\alpha$ over the unit torus $x \in [0,1)$. For an integer $n \geq 2$, let $[n] = \{1,2,3,\ldots,n\}$ denote the set containing the first $n$ natural numbers (we will use it with $n = 24$ for our purposes). We then define the *skew* product $([0,1)\times[24],T_\alpha)$ as

$$T_\alpha(y,m) = (R_\alpha y, \mathbf{P}_\alpha(y)m), \quad m \in [24], \tag{10.6}$$

where $\mathbf{P}_\alpha(x)$ is defined as in 10.2. We recall that $\mathbf{P}_\alpha(x)$ is defined by the interval exchange transformation $T_\alpha$, and in particular is such that $\mathbf{P}_\alpha(x)$ is constant on the two special intervals $[0,1-\alpha)$ and $[1-\alpha,1)$.

To prove Step 1, we will show that $([0,1),R_\alpha)$ is uniquely ergodic. Assume the contrary; that is, there exists a Borel measure $\nu(.)$, defined on the Borel sets of [0,1), which is $R_\alpha$-invariant and $\nu \neq \lambda$. Since Borel sets are generated by intervals, this implies there is an interval $I_0 \subset [0,1)$ such that $\nu(I_0) \neq \lambda(I_0) = \text{length}(I_0)$.

We can assume even that $\nu$ is $R_\alpha$-ergodic. Indeed, the space of Borel probability measure on the compact set unit torus [0,1) (again, taken as a circle so it is compact) is itself compact. (As we will later use the compactness of the space of measure several times, we note here that it is true for any underlying compact set; the standard proof is based upon the Riesz Representation Theorem.) If $T : [0,1) \to [0,1)$ is Borel measurable then the subset of $T$-invariant Borel probability measures on $[0,1)$ is closed. Thus it is compact, and clearly is convex, and so the Krein-Milman theorem implies that it is spanned by extremal points [6]. The extremal points are exactly the $T$-ergodic measures, and we will select $\nu$ to be such an extremal point with $T = R_\alpha$.

Birkhoff's ergodic theorem tells us that, for almost every (with respect to $\nu$) $y \in [0,1)$

$$\lim_{k\to\infty} \frac{1}{K} \sum_{\substack{1\leq k\leq K: \\ y+k\alpha\in I_0 \ (mod\ 1)}} 1 = \nu(I_0). \tag{10.7}$$

However, the classical result of uniform distribution tells us that

$$\lim_{k\to\infty}\frac{1}{K}\sum_{\substack{1\le k\le K:\\ y+k\alpha\in I_0 \ (mod\ 1)}} 1 = \text{length}(I_0) \tag{10.8}$$

for every $y \in [0,1)$ and every irrational number $\alpha$. Since $I_0$ was selected such that $\nu(I_0) \neq \text{length}(I_0)$, 10.8 contradicts 10.7, and thus our assumption that such a measure $\nu$ exists is false. Therefore $([0,1),R_\alpha)$ is uniquely ergodic. (For alternative proof that $([0,1),R_\alpha)$ is uniquely ergodic, see Theorem 3.12 in Furstenberg [6].)

We are now ready to prove Step 1, which is equivalent to the statement that the skew product $([0,1)\times[24],T_\alpha)$ (as defined in 10.6) is uniquely ergodic (noting that the $\sigma$-algebra is always Borel sets). Again we assume the contrary, that this is a Borel measure $\nu(.)$ (defined on the Borel sets of $[0,1)\times[24]$) which is $T_\alpha$-invariant and $\nu \neq \lambda\times\mu_{24}$, where the second term in the product measure, $\mu_{24}$, is the normalized counting measure, i.e. $\mu_{24} = \frac{1}{24}$ cardinality.

It follows that there must be a continuous function $f_0$ on $X = [0,1)\times[24]$ such that

$$\int_X f_0\, d\nu \neq \int_X f_0\, d\lambda\times\mu_{24}. \tag{10.9}$$

By repeating the argument above and again applying the Krein-Milman theorem, we can again assume the $\nu$ is $T_\alpha$-ergodic.

As $(X,\lambda_{24},T_\alpha)$ with $X=[0,1)\times[24]$ is ergodic (see Lemma 17), Birkhoff's ergodic theorem tells us that for every continuous function $f$ on $X$, and almost every (with respect to $\lambda$) $y\in[0,1)$

$$\lim_{k\to\infty}\frac{1}{K}\sum_{k=1}^{K-1} f(T_\alpha^k z) = \int_X d\lambda\times\mu_{24} \tag{10.10}$$

holds for every $z$ of the form $z=(y,m)$, $m\in[24]$.

Repeating the same argument as previously with measure $\nu$ instead of $\lambda_{24}$,
Birkhoff's ergodic theorem gives us the analogous result, for almost every (with respect to $\nu$) $x\in X$

$$\lim_{k\to\infty}\frac{1}{K}\sum_{k=1}^{K-1} f(T_\alpha^k x) = \int_X d\nu \tag{10.11}$$

holds for every continuous function $f$ on $X$. By combining 10.10 and 10.11, and using a function $f$ satisfying the condition in 10.9, we obtain that the set of $x \in X$ satisfying 10.11 is contained in a product set $A_1 \times [n]$ such that $A_1 \subset [0,1)$, with $\lambda(A_1) = 0$. Let $\nu_1(.)$ denote the projection of $\nu(.)$ onto $[0,1)$ (in the obvious way, i.e. $\nu_1(B) = \nu(B \times [24])$ for every Borel set $B \subset [0,1)$). Since $\nu_1$ is $R_\alpha$-invariant and $([0,1), R_\alpha)$ is uniquely ergodic, as proved above, we conclude that $\nu_1 = \lambda$. But this is impossible as $\lambda(A_1) = 0 \neq 1 = \nu_1(A_1)$. This is a contradiction and thus again our assumption of the existence of such a measure $\nu$ is false. This proves Step 1.

We define the following notation: we say that $x \in [0,1)$ is a $\alpha$-legitimate point if $x \neq \{-h\alpha\}$, $h \geq 0$ integer, i.e., if $T_\alpha^k x$ is well defined for every integer $k \geq 0$.

**Step 2:** *For every continuous function $f$ on $X = [0,1) \times [24]$*

$$\frac{1}{K} \sum_{k=0}^{K-1} f(T_\alpha^k z) \to \int_X f \, d\lambda \times \mu_{24} \text{ as } K \to \infty \tag{10.12}$$

*uniformly for all $z = (x,m)$ where $m \in [24]$ and $x \in [0,1)$ is $\alpha$-legitimate.*

We prove Step 2 by contradiction. If the uniform convergence in 10.12 fails, then there exists some continuous function $f_0$ and an $\varepsilon > 0$ such that for each $N$ there exists $n > N$ and exists a $z_n(x_n, j)$ with $j = j(n) \in [24]$ and an $\alpha$-legitimate $x_n \in [0,1)$ such that

$$\left| \frac{1}{K} \sum_{k=0}^{K-1} f_0(T_\alpha^k z) - \int_X f_0 \, d\lambda \times \mu_{24} \right| \geq \varepsilon. \tag{10.13}$$

We define the normalized "counting measure" (as a probability measure)

$$\nu_n = \frac{1}{n} \sum_{i=0}^{n-1} \delta(T_\alpha^i z_n), \tag{10.14}$$

where for every Borel set $B \subset X$, $\delta(y)(B) = 1$ or 0 accordingly as $y \in B$ or $y \neq B$, respectively. By using 10.14 we can rewrite the equation in 10.13 in the form

$$\left| \int_X f_0 \, d\nu_n - \in_X f_0 \, d\lambda \times \mu_{24} \right| \geq \varepsilon. \tag{10.15}$$

Again we use the crucial fact that the set of Borel probabilities on the compact set $X$ is itself too compact, and thus there is some convergent sub-sequence $\nu n_i$ of measures, which converges to a Borel probability measure $\nu$ on $X$. More precisely, $\nu_{n_i} \to \nu$ as $i \to \infty$. It follows from 10.14 that $\nu$ is $T_\alpha$-invariant. Since $([0,1) \times [24], T_\alpha)$ is uniquely ergodic (as shown in Step 1), we can conclude that $\nu = \lambda \times \mu_{24}$. But $\nu_{n_i} \to \lambda \times \mu_{24}$ contradicts 10.15. This contradiction proves our assumption false and thus proves Step 2. Therefore we get the following upgrade of Theorem 1:

**Theorem 9** (**upgrade to Theorem 1**). *For every irrational slope, geodesic $L(t)$, $0 \leq t < \infty$, $t \to \infty$, with the given slope is uniformly distributed on the cube surface in the usual Weyl sense (see Lemma 1) for every non-pathological starting point $L(0)$.*

□

# CHAPTER 11

# GENERALIZATIONS

We explore three generalizations of the results in Theorem 1, 2, and 9: (1) geodesic flow on the platonic solids beyond the cube, (2) billiards path in polyominoes, and (3) geodesic flow on the surfaces of solid polyominoes.

Intuitively, *polyominoes* are "finite connected square lattice regions" in the plane. More precisely, polyominoes are shapes constructed by connecting a finite number of unit sized squares, which we will refer to as *(component) unit squares*, each joined together with at least one other square. Polyominoes may have holes in them, but they must be $XY$-connected. By $XY$-connected, we mean "square connected": or, given any two unit squares in the polyomino, a particle starting in one square could travel to the second square by repeatedly moving to adjacent unit squares which share an edge. We note all steps in this process must be travel through an edge shared by two unit squares, meaning traveling horizontally (parallel to the $X$-axis) or vertically (parallel to the $Y$-axis), but never diagonally.

The concept of a solid polyomino is a straightforward 3-dimensional generalization of the 2-dimensional polyomino described above. Intuitively, solid polyominoes are "finite connected 3-dimensional cube lattice regions". More precisely, a solid polyomino is made by connecting a finite number of unit cubes, which we will similarly refer to as *(component) unit cubes*, each joined together with at least one other cube along a square face. The surface of a solid polyomino need not be connected, in the case of a "hollow solid"; for example a solid polyomino which is located between two closed surfaces, one contained within the other. They must, however, be $XYZ$-connected. Being $XYZ$-connected means, given any two unit cubes in the solid

polyomino, a particle starting in one of the cubes can travel to the second cube by repeatedly moving to adjacent unit cubes which share a face. We note all steps in this process must travel through a face shared by two unit cubes, meaning traveling parallel to the $X$-axis, parallel to the $Y$-axis, or parallel to the $Z$-axis, but never diagonally,

The concept of polyomino (and solid polyomino) became well accepted after the publication of the book of Solomon W. Golomb with the same title [7].

We carry out the detailed discussion of these three generalizations in reverse order.

**1. Geodesic flow on the surfaces of solid polyominoes.** Let $\mathbf{U}$ be a solid polyomino, and $\partial\mathbf{U}$ be its boundary surface. The boundary $\partial\mathbf{U}$ must be closed and orientable, but it is not necessarily connected. Let $\Delta \subseteq \partial\mathbf{U}$ be a connected component of $\partial\mathbf{U}$. We refer to $\Delta$ as a "Lego surface". Then $\Delta$ is a finite union of unit squares satisfying the following properties: (1) each pair of (component) unit squares is either disjoint or their intersection is a common vertex or a common edge, (2) for any two different (component) unit squares, $\mathscr{U}, \mathscr{U}'$ this is a finite sequence of component unit squares $\mathscr{U}_1, \mathscr{U}_2, \ldots, \mathscr{U}_n$ such that $\mathscr{U}_i$ and $\mathscr{U}_{i+1}$ share a common edge, for $0 \leq i \leq n$, and $\mathscr{U}_0 = \mathscr{U}$, $\mathscr{U}_{n+1} = \mathscr{U}'$, (3) $\Delta$ is a closed surface: i.e. exactly two unit squares share a common edge, and finally, (4) $\Delta$ is orientable: it is possible to orient the boundaries of the component unit squares such that when two unit squares meet, the orientations along the shared edge run in opposite directions.

Let $\mathscr{E}$ be an edge of a component unit square of the Lego surface $\Delta$. We basically repeat the definitions/arguments made in Chapter 2 to reformulate the problem of a geodesic on $\Delta$ to a discrete problem. Consider a geodesic on the surface $\Delta$ that starts from edge $\mathscr{E}$ with angle $0 < \theta < \pi/2$, such that the slope $\tan\theta$ is irrational. By symmetry given from rotating the solid polyomino $\mathbf{U}$ through 3-space we may also assume the $\pi/4 < \theta < \pi/2$, implying $\tan\theta > 1$. The geodesic returns/crosses the same edge $\mathscr{E}$ infinitely often with angles $\theta$ or $\theta + \pi/2$. We chose to just focus on the $\theta$-crossings.

Just like in Chapter 2, **directed $\theta$ edge** means the following: assume a particle is moving with constant speed on a geodesic which enters a component unit square $\mathscr{U}$ of $\Delta$ by crossing edge $\mathscr{E}$, and let $\mathscr{E}$ be the bottom edge of the square $\mathscr{U}$. The positive (counter-clockwise) orientation of $\mathscr{U}$ induces an orientation $\vec{\mathscr{E}}$ of the edge $\mathscr{E}$, and the local part of the geodesic entering

$\mathscr{U}$ is a directed line segment. This directed line segment and directed edge $\vec{\mathscr{E}}$ with its induced orientation, determine an angle, which we require to be $\theta$ (enforced through orientation of $\vec{\mathscr{E}}$).

To give this a quantitative description, it is convenient to assume that, without loss of generality, $\mathscr{E}$ is the bottom edge of the unit square $\mathscr{U}$. We then identify the edge $\mathscr{E}$ with unit interval $[0,1)$ by using an orientation $\vec{\mathscr{E}}$ of $\mathscr{E}$. The directed $\theta$ edge geodesic flow on Lego surface $\Delta$ moves the left endpoint of $\vec{\mathscr{E}}$ (i.e. 0) to $\alpha = 1/\tan\theta$ $(0 < \alpha < 1)$ on the corresponding directed edge above $\mathscr{E}$ in the unit square $\mathscr{U}$, and similarly moves $1-\alpha$ on $\vec{\mathscr{E}}$ to 1 on the edge above it, and moves 1 on $\vec{\mathscr{E}}$ to $\alpha$ on the neighboring directed edge in the neighboring component unit square. More precisely, the open interval $(0,1-\alpha)$ in $\vec{\mathscr{E}}$ is mapped to the open interval $(\alpha,1)$ on the directed edge $\vec{\mathscr{E}}'$ above $\vec{\mathscr{E}}$, and $(1-\alpha,1)$ in $\vec{\mathscr{E}}$ is mapped to the $(0,\alpha)$ part on the directed edge $\vec{\mathscr{E}}''$ neighboring $\vec{\mathscr{E}}$'. This essentially is an $\alpha$-shift $x \to x+\alpha$ (modulo 1), or formally,

$$\vec{\mathscr{E}}(0,1-\alpha) \to \vec{\mathscr{E}}'(\alpha,1) \text{ and } \vec{\mathscr{E}}(1-\alpha,1) \to \vec{\mathscr{E}}''(0,\alpha), \tag{11.1}$$

noting that 11.1 is a perfect analog of 2.1. Exactly the same way as we extended 2.1 over all 24 directed edges of the unit cube surface, we can extend 11.1 over all directed edges of Lego surface $\Delta$ (of course each edge is taken with both possible orientations).

Let $N$ be the total number of edges $\mathscr{E}$ of the component unit squares on the Lego surface $\Delta$ (note that although every edge $\mathscr{E}$ is contained in two squares, here we only count it with multiplicity only one). As every edge has two orientations, Lego surface $\Delta$ will have $2N$ directed edges.

Again mirroring Chapter 2, where we glued the 24 directed edges of the cube together to form the interval $[0,24)$, here we glue the $2N$ directed edges of $\Delta$ together to form the interval $[0,2N)$. Then applying 11.1 to every directed edge (switching to half open intervals) induces an interval exchange transformation $T = T(\Delta)$. $T$ translates a half open interval of the form $[r-1, r-\alpha)$, for $r$ an integer, $1 \le r \le 24$, to some other half-open interval of the form $[r'-1+\alpha, r')$, for $r'$ an integer with $1 \le r' \le 24$, and translates half-open intervals of the form $[r-\alpha, r)$ to half-open intervals of the form $[r''-1, r''-1+\alpha)$, once again with $r''$ an integer, $1 \le r'' \le 24$. The mapping $T$ has discontinuous jumps at the points $x$, where the fractional part of $x$, $\{x\} = 1-\alpha$ or, $\{x\} = 0$. We again call these jumps the *singular* points of $T$. And again

we note the fact that $T$ is a Lebesgue measure preserving transformation.

The key characteristic of the transformation $T$ is the slope of the geodesic inducing the irrational shift, $0 < \alpha < 1$, so we denote the shift $T = T(\Delta) = T_\alpha(\Delta)$. Again we note this shift can similarly be done on any irrational slope, not just $0 < \alpha < 1$, which was just a result of our simplifying assumption that using symmetry we can use take $\pi/4 < \theta < \pi/2$. Notably, a transformation can similarly be constructed on the $\theta^\perp$-edge crossings of the geodesic; we denote this transformation $\widetilde{T} = \widetilde{T}_{-1/\alpha}$.

Again the main difficulty is to prove the analog of Lemma 17: $T_\alpha(\Delta)$ (and similarly $\widetilde{T}_{-1/\alpha}(\Delta)$) are ergodic if $\alpha$ is an irrational real number. Repeating the proof of Lemma 7 with the trivial change that $[0,24)$ is replaced by $[0,2N)$, we can easily prove ergodicity, with the exception of one caveat: Lemma 16. While the connectivity of a concrete relatively small graph of 24 vertices and 48 edges can be proves by brute force checking, the analogous statement for an entire class of Lego surfaces requires checking connectivity of infinitely many graphs, which certainly cannot be done by brute force. So, we introduce a generalization of Lemma 16 and a "clever" proof of it (which will include an explicit proof of Lemma 16 as well).

As the directed cube-surface-reachability graph $\mathbf{G}_{cube}(\alpha)$ was defined by the transformation $T_\alpha(cube)$ via rule 2.1, we analogously define for every Lego surface $\Delta$ its directed $\Delta$-reachability graph $\bar{\mathbf{G}}_\Delta(\alpha)$ by the transformation $T_\alpha(\Delta)$ via rule 11.1, and then again by ignoring orientation on this directed graph we obtain the
$\Delta$-reachability graph $\mathbf{G}_\Delta(\alpha)$. We prove the following generalization of Lemma 16:

**Lemma 18.** *If $\alpha$ is irrational, then for every Lego surface $\Delta$ the $\Delta$-reachability graph $\boldsymbol{G}_\Delta(\alpha)$ is connected.*

As usual let $0 < \alpha < 1$. Lego surface $\Delta$ must contain a vertex which is a corner (vertex of degree 3) of the underlying solid polyomino $\mathbf{U}$ (i.e. $\Delta \subseteq \partial\mathbf{U}$). We will label the 8 vertices and the three squares that form the corner. Label the corners $A, B, \ldots, H$, with $F$ being the corner itself, and the three squares $S_1 = ABFE$, $S_2 = BCGF$, and $S_3 = EFGH$; see Figure 11.1. For notational convenience, when referring to edges, we assume the order of the vertices is the direction induced on the edge (i.e. edge $AB$ and edge $BA$ are the same edge with opposite orientations).

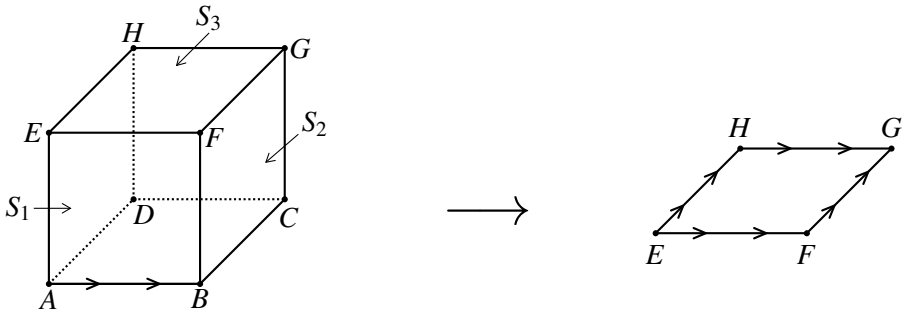


Figure 11.1: The direction of the line leaving $AB$ induces a direction on the edges of the square $EFGH$.

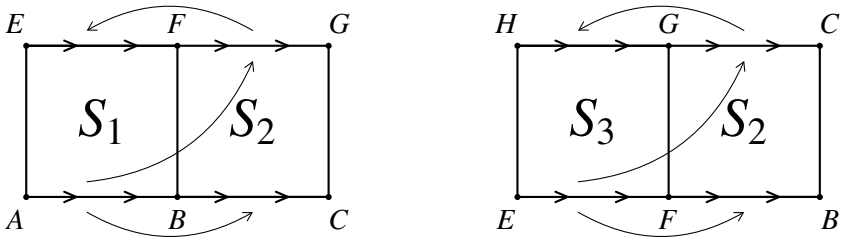


Figure 11.2: Following the direction of the line allows us to extend the induced directions from one square to an adjacent one, in both the forwards and backwards directions.

Suppose a geodesic is coming from edge $AB$ and is moving towards the corner point $F$. Then from this orientation $AB$, by applying 11.1, it is connected to two orientations of the corner, the shared edge between $S_1$ and $S_2$, $EF$, and the shared edge between $S_2$ and $S_3$, $FG$, (i.e. the directed edge above it and the directed edge neighboring that one), see Figure 11.2. By repeated application of the rule in 11.1 (and as necessary applying it in reverse), we can see that the orientation $AB$ is connected to an orientation of each of the other eight edges within the three squares that form this corner: using $EF$ as a base as we did with $AB$ we get $HG$ and $GC$; using $FG$ as a base we get $BC$ by using 11.1 and $EH$ by using 4.1 in reverse. Finally, using reverse 11.1 with $GC$ as a base gives both $FB$ and $EA$. See Figure 11.3.

Moreover, using this rule, for any component square $U$ in $\Delta$ that borders $S_3$, we see there is an orientation for each edge of $U$ which is connected to the directed edges induced by the orientations above. In fact, the orientations above will extend in a repeating pattern; parallel edges will be oriented the same way, see Figure 11.4.

Since starting from $S_3$ and moving to a neighbor in each step we can reach every component unit square in $\Delta$, we have that at least one orientation of every edge in $\Delta$ is connected to the

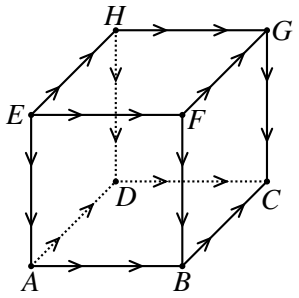


Figure 11.3: This in turn induces a direction on all edges of the cube.

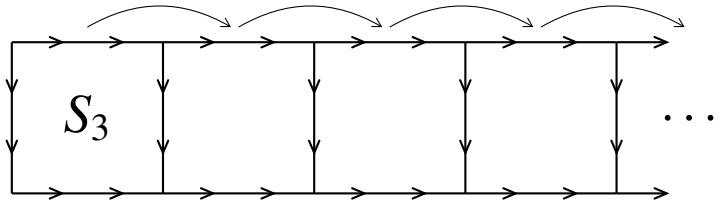


Figure 11.4: In the case of general polyominoes, this can be extended to long chains of adjacent squares.

original directed edge $AB$. This tells us that if the $\Delta$-reachability graph $\mathbf{G}_\Delta(\alpha)$ is disconnected, it must have exactly two components (say) $Comp_1$ and $Comp_2$ which form an "exact split", that is, one orientation of each edge is in $Comp_1$, and the other orientation of each edge is in $Comp_2$. We will next show that such an "exact split" is impossible.

We assume that $AB \in Comp_1$. Then $Comp_1$ consists of exactly the directed edges reachable from $AB$ (the ones described above), Then the opposite orientation of each directed edge in $Comp_1$ is in $Comp_2$ (e.g. $BA \in Comp_2$). But as above, this means that $\{BF, AE, HE\} \subset Comp_2$. However using $BF$ as a base, 11.1 gives us $EH \in Comp_2$. But $EH \in Comp_1$ as above, so this implies $Comp_1 = Comp_3$. Thus we conclude that the "exact split" was impossible, so the entire graph $\mathbf{G}_\Delta(\alpha)$ must be connected for $0 < \alpha < 1$.

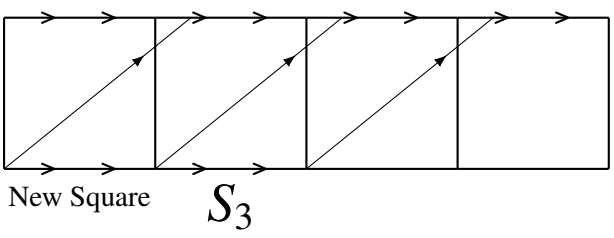


Figure 11.5: The method of inducing directions on edges can be generalized for the case of $|\alpha| < 1$ by including an extra square and "pulling back" the shift of induced directions.

We finally note that while we did the proof above for $0 < \alpha < 1$, it works equally well for all values of $\alpha$, by working a square over from where we want, we can see that proof is equivalent for $\alpha$ modulo 1, see Figure 11.5. Moreover, by redefining our directed edges to be the $\theta^{\perp}$ directed edges, we can enforce $0 < \alpha < 1$, and connectivity of the directed $\theta$ edges follows clearly from the connectivity of the directed $\theta^{\perp}$ edges. This completes the proof of Lemma 18. □

We further note that this provides a proper proof of Lemma 16, as a cube is a solid polyomino with one component.

Repeating the proof of Lemma 17 with replacing 24 with $2N$, replacing Lemma 16 with Lemma 18, and repeating the arguments of "upgrading by unique ergodicity" as in Theorem 9, we obtain a generalization of Theorem 1, 2 and 9.

**Theorem 10.** *(i) Let $\Delta$ be an arbitrary Lego surface. For every irrational slope, geodesic $L(t)$, $0 \leq t < \infty$, $t \to \infty$ of given slope is uniformly distributed on closed surface $\Delta$ in the usual Weyl sense for every non-pathological starting point $L(0)$.*

*(ii) Also, geodesic $L(t)$, $0 \leq t < \infty$, $t \to \infty$ is uniformly distributed on a closed surface $\Delta$ in the Birkhoff sense for almost every starting point $L(0)$.*

□

**Remark.** We emphasize the fact that in the proof we only used the following properties of Lego surface $\Delta$: (1) it is orientable, (2) it is closed, and (3) it consists of unit square components that are "square connected".

**2. Billiard path in polyominoes.** The following argument works for any polyomino, but for simplicity we consider a concrete polyomino: let $L$ be the L-shaped polyomino with three component squares, see Figure 11.6. We previously discussed how the geometric trick of *unfolding* reduces any piecewise linear billiard path in a unit square into a torus line on a $2 \times 2$ square (see Chapter 1). Using the same ideas, we will show that any billiard path on the L-shaped table $\mathbf{L}$ can be reduced to a geodesic on an orientable closed flat surface $\mathbf{L}^{(2)}$, where $\mathbf{L}^{(2)}$ is the closed surface obtained by taking $2 \times 2 = 4$ rotated copies of versions of $\mathbf{L}$ and

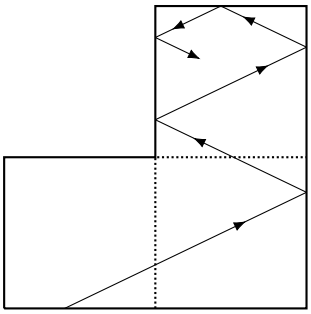

Figure 11.6: A billiard path on an L-shaped polyomino. Just like billiard paths in the unit square, when the path hits an edge it reflects with angle of reflection equal to angle of incidence.

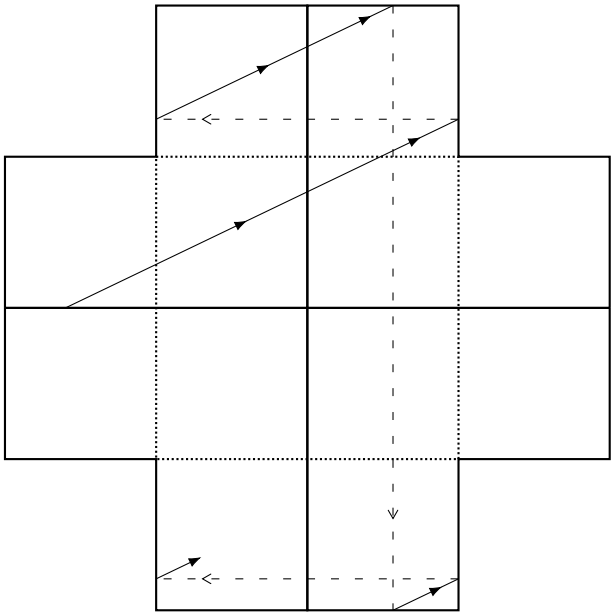

Figure 11.7: Just like the unit square, billiard paths on the L-shaped polyomino can be unfolded though reflection and is equivalent to a torus line on the reflected figure.

gluing them together in a particular way, see Figure 11.7.

Let $\mathbf{L}_k$, $k = 0,1,2,3$ be these four copies of $\mathbf{L}$, each rotated by an angle of $k\pi/2$ respectively. We glue the $\mathbf{L}_k$ together along the long edges and obtain a 12-gon polyomino consisting of 12 component unit squares. We then make a pairwise identification of the boundary edges of the 12-gon, such that each edge is paired to the one perpendicularly across from it (that is, if you drew a line perpendicular to a boundary edge, the two boundary edges it intersects are paired). This induces an orientable closed flat surface $\mathbf{L}^{(2)}$.

A straightforward geometric consideration shows that unfolding a piecewise linear billiard path on $\mathbf{L}$ is equivalent to a geodesic on the orientable closed flat surface $\mathbf{L}^{(2)}$. We note that $\mathbf{L}^{(2)}$

consists of unit square components which are "square-connected". Thus, the three conditions mentioned in the Remark following Theorem 10 are satisfied, and we can repeat the arguments used in Theorem 10.

The argument in this case is actually a bit simpler than in Theorem 10, as we do not need to consider both orientations of each edge; rather it is sufficient to just have the synchronized orientation for vertical edges and the synchronized orientation for horizontal edges (edges here including all edges, not just boundary edges). This means that the underlying interval in the analog of Lemma 17 will be $[0,24)$ rather than $[0,48)$, and we do not need any analog of Lemma 18, as connectivity will follow directly from the square connectivity of the polyominoes. Thus we obtain the following analog of Theorem 10.

**Theorem 11.** *(i) Let* $\boldsymbol{L}$ *be the L-shaped polyomino with three component unit squares. For every irrational slope, billiard path* $\boldsymbol{x}(t)$, $0 \leq t < \infty$, $t \to \infty$ *of given initial slope is uniformly distributed in* $\boldsymbol{L}$ *in the usual Weyl sense for every non-pathological starting point* $\boldsymbol{x}(0)$.

*(ii) Also, billiard path* $\boldsymbol{x(t)}$, $0 \leq t < \infty$, $t \to \infty$ *is uniformly distributed in L in the Birkhoff sense for almost every starting point* $\boldsymbol{x}(0)$.

□

We again note that Theorem 11 can be extended to any polyomino using the same argument.

**3. Geodesic flow on the platonic solids beyond the cube.** The case of the tetrahedron is quite simple, because the flat surface of a regular tetrahedron tiles the plane. Using this fact, analogs of Theorem 1, 2, and 9 were already proved in Beck [2]. Beck [2] in fact settled the more general case of **equifacial tetrahedrons** where the four faces are any arbitrary congruent acute triangles.

The regular octahedron and icosahedron both have regular triangle faces. Consider two faces that share an edge. We rotate through the edge to make the triangles coplanar, and form a 60-120 degree rhombus, see Figure 11.8. The only substantive alteration to the arguments for Theorem 1, 2, and 9, is altering the way in which the interval exchange function $T$ is defined

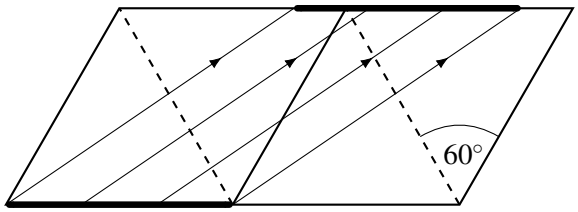


Figure 11.8: By unfolding pairs of equilateral triangle sides, the octahedron and icosahedron can be interpreted as having 60-120 rhomboidal sides. The argument for lines on the cube follows for these shapes with rhombus sides, assuming that the slope is 60-degree-irrational.

(see 2.1), now being on a rhombus instead of a square. Treating the faces as rhombuses, however, well defines the interval exchange from one face to another.

We need a couple other small alterations to the proof. For one, the surface-reachability graph in Lemma 16 needs to be altered, but we are again left with a concrete, relatively small graph that can be proved by brute force checking.

Second, "irrational slope" must be changed to "60-degree-irrational slope",
which is defined as the following. Consider a regular triangle lattice $\mathscr{L}$ on the plane. We say a straight line has "60-degree-rational slope" if there is a parallel line to it which contains at least two lattice points of $\mathscr{L}$; otherwise we say the slope is "60-degree-irrational".

With these particulars resolved, the argument for Lemma 17 follows exactly for these cases as well and we get the following theorem.

**Theorem 12.** *(i) For every 60-degree-irrational slope, geodesic $L(t)$, $0 \leq t < \infty$, $t \to \infty$ of given slope is uniformly distributed on both the (regular) octahedron and icosahedron surfaces in the usual Weyl sense for every non-pathological starting point $L(0)$.*

*(ii) Also, geodesic $L(t)$, $0 \leq t\infty$, $t \to \infty$ is uniformly distributed on both the (regular) octahedron and icosahedron surfaces in the Birkhoff sense for almost every starting point $L(0)$.*

□

# CHAPTER 12

# OPEN QUESTIONS

Many open questions regarding cube lines still remain. Firstly, quantitative results for cube lines with arbitrary irrational slope remains open, as our quantitative results only work for a select set of "special" slopes. Moreover, even these partial results for "special" slopes do not extend to geodesics over more general figures, such as the $2 \times 1 \times 1$ rectangular prism. Additionally, as mentioned in Chapter 11, our results for cube lines can be extended to the octahedron and icosahedron (and the case of the tetrahedron being previously solved), but the case of a geodesic on the final platonic solid, the dodecahedron, remains open.

# REFERENCES


[1] Beatty: *Problem 3173*, American Mathematical Monthly 1926, p 159.

[2] Beck: *From Khinchin's conjecture on strong uniformity to superuniform motions*, Mathematika 2015, p 591-707.

[3] Birkhoff: *Proof of the ergodic theorem*, Proceedings of the National Academy of Sciences 1931, p 656-660.

[4] Carathéodory: *Über den Wiedenkehrsatz von Poncaré*, Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin 1919, p 296-301.

[5] Erdős, Turán, *On a problem in the theory of uniform distribution I & II*, Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen 1948 p 27-28, 49-52.

[6] Furstenberg: *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University press 1981

[7] Golomb: *Polyominoes: Puzzles, Patterns, Problems, and Packings*, Princeton University press 1996

[8] Gutkin: *Billiards on almost integrable polyhedral surfaces*, Ergodic Theory and Dynamical Systems, 1984, p 569-584.

[9] Hardy, Littlewood: *Some problems of Diophantine approximation*, Acta Mathematica 1914, p 193-239.

[10] Hardy, Wright: *An introduction to the Theory of Numbers*, Oxford University press 1979

[11] Hellakalek: *Random and Quasi-Random Point Sets*, Springer 1998.

[12] Katok, Zemlyakov: *Topological transitivity of billiards in polygons*, Mathematical Notes 1975, p 291-300.

[13] Koksma: *Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1*, Mathematica B 1942/43, p 7-11.

[14] Konig, Szucs: *Mouvement d'un point abondonne a l'interieur d'un cube*, Rendiconti del Circolo Matematico di Palermo 1913, p 79-90.

[15] Kuipers, Niederreiter: *Uniform distribution of sequences*, John Wiley 1974.

[16] Lebesgue: *Leçons sur l'Intégration et la Recherche des Fonctions Primitives*, Gauthier Villars 1904.

[17] Ostrowski: Bemerkungen zur Theorie der diophantischen Approximationen, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg 1921, p 77-98.

[18] Poincaré: *Sur le problème des trois corps et les équations de la dynamique*, Acta Mathematica 1890, p 262-490.

[19] Rayleigh: *The Theory of Sound* Macmillan 1894.

[20] Weyl: *Über die Gleichverteilung von Zahlen mod. Eins*, Mathematische Annalen 1916, p 313-352.

[21] Wilkinson: *Ergodic properties of a class of piecewise linear transformations*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 1975, p 303-328.