

# ERGODIC THEORY AND CONTINUED FRACTIONS

PAUL GUTKOVICH

ABSTRACT. We explore the connections between ergodic theory and the theory of continued fractions. After developing the necessary measure-theoretic and ergodic theory background, we establish that the Gauss map  $T(x) = \{1/x\}$  is ergodic with respect to the Gauss measure  $d\mu = \frac{1}{\ln 2} \frac{dx}{1+x}$  on  $(0, 1)$ . This allows us to apply Birkhoff's Ergodic Theorem to prove the Gauss–Kuzmin theorem: for almost every real number, the digit  $k$  appears in its continued fraction expansion with asymptotic frequency  $\log_2\left(1 + \frac{1}{k(k+2)}\right)$ . We conclude by deriving Khinchin's constant  $K_0 \approx 2.685$ , the universal geometric mean of continued fraction digits.

## CONTENTS

1. Introduction	1
2. Convergents	2
3. Ergodic Theory	4
4. Ergodicity of the Gauss Map	5
5. The Gauss-Kuzmin Theorem	9
5.1. Pointwise Gauss-Kuzmin Theorem	10
5.2. Original Gauss-Kuzmin Theorem	11
6. Khinchin's Constant	11
References	12

## 1. INTRODUCTION

A simple continued fraction is an object of the form

$$a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}$$

where all the  $a_i$  are in  $\mathbb{N}$ . This expression could have finitely many terms, in which case we say that the simple continued fraction terminates. For brevity, we will refer to simple continued fractions just as continued fractions. We often denote such a continued fraction as  $[a_0; a_1, a_2, \dots]$ , or  $[a_0; a_1, a_2, \dots, a_n]$  if it terminates. In the case where  $a_0 = 0$ , we will often drop the first term and just write  $[a_1, a_2, \dots]$ .

DEFINITION 1.1 (Convergents). Given a continued fraction defined by a sequence of  $a_i$ , we define its  $n$ th convergent as

$$C_n = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_n}}}$$

---

*Date:* May 8, 2026.

A continued fraction is said to converge to a real number  $x$  if and only if its sequence of convergents  $C_0, C_1, \dots$  converges to  $x$ . It is well known that every positive real number admits a unique continued fraction expansion, which terminates if and only if the number is rational (see [Kar13, Theorem 1.17]). This raises the natural problem of explicitly determining the continued fraction expansion of a given real number. To address this, we introduce the Gauss map.

**DEFINITION 1.2 (Gauss Map).** Define the Gauss map  $T : (0, 1) \rightarrow [0, 1)$  with  $T(x) = \{1/x\}$ , where  $\{r\}$  denotes the fractional part of  $r$ .

Consider a real number  $x \in (0, 1)$  with continued fraction representation  $[a_1, a_2, a_3, \dots]$ , i.e.

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}.$$

Then

$$\begin{aligned} T(x) = \{1/x\} &= \left\{ a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}} \right\} \\ &= \frac{1}{a_2 + \frac{1}{a_3 + \dots}} = [a_2, a_3, \dots]. \end{aligned}$$

Thus, the Gauss map has the property that it shifts terms to the left in the continued fraction expansion of a number. Using this, we have that  $a_i = \lfloor 1/T^{i-1}(x) \rfloor$ , so the Gauss map allows us to find the terms of a number's continued fraction expansion.

A central question is to understand the statistical behavior of the continued fraction expansion of a "typical" real number. In particular, we ask: how frequently does a given digit  $k$  appear in the continued fraction expansion  $[a_0; a_1, a_2, \dots]$  of a randomly chosen real number? And what can be said about the growth rate of those digits?

This paper answers both questions using tools from ergodic theory. The key insight is that the Gauss map is an ergodic transformation with respect to a natural probability measure on  $(0, 1)$ .

The paper is organized as follows. In Section 2, we establish two properties of convergents needed to study the inverse of the Gauss map in later sections.

Section 3 introduces the necessary background from ergodic theory, including measure-preserving transformations, ergodicity, and Birkhoff's Ergodic Theorem.

In Section 4, we establish that the Gauss map is ergodic with respect to the Gauss measure  $\mu(S) = \frac{1}{\ln 2} \int_S \frac{dx}{1+x}$ .

Section 5 applies this to prove the Gauss–Kuzmin theorem, which states that for almost every  $x \in (0, 1)$ , the digit  $k$  appears in the continued fraction of  $x$  with limiting frequency  $P(k) = \log_2 \left( 1 + \frac{1}{k(k+2)} \right)$ .

Section 6 derives Khinchin's constant: for almost every real  $x$ , the geometric mean of the continued fraction digits satisfies  $\lim_{n \rightarrow \infty} (a_1 a_2 \cdots a_n)^{1/n} = K_0 \approx 2.685$ .

## 2. CONVERGENTS

Before developing the ergodic theory machinery, we establish two fundamental properties of the convergents of a continued fraction. The first gives a recursive

formula for computing them, and the second is an identity relating consecutive convergents that will be used in Section 4 when analyzing the Gauss map's inverse branches.

**PROPOSITION 2.1.** *For any continued fraction  $[a_0; a_1, \dots]$ , define two sequences recursively with  $p_{-1} = 1, q_{-1} = 0, p_0 = a_0, q_0 = 1$ , and*

$$p_k = a_k p_{k-1} + p_{k-2}, \quad q_k = a_k q_{k-1} + q_{k-2}.$$

*Then  $p_k/q_k$  equals the  $k$ th convergent  $C_k$  for all  $k \geq 0$ .*

*Proof.* We will proceed by induction on  $k$ . The  $k = 0$  case is trivial. For  $k \geq 0$ , we have the following equality for two convergents:

$$[a_0; a_1, a_2, \dots, a_{k+1}] = [a_0; a_1, \dots, a_{k-1}, a_k + 1/a_{k+1}].$$

Let  $p_i, q_i$  denote the recursive sequences defined for  $[a_0; a_1, a_2, \dots, a_{k+1}]$ , and  $p'_i, q'_i$  denote the recursive sequences for  $[a_0; a_1, \dots, a_{k-1}, a_k + 1/a_{k+1}]$ . For all  $i \leq k-1$ ,  $p_i = p'_i, q_i = q'_i$ . By the inductive hypothesis, we also have

$$p'_k/q'_k = [a_0; a_1, \dots, a_{k-1}, a_k + 1/a_{k+1}]$$

From the equality relating the convergents, this gives us

$$p'_k/q'_k = [a_0; a_1, \dots, a_{k+1}].$$

We have that

$$\begin{aligned} [a_0; a_1, \dots, a_{k+1}] &= \frac{p'_k}{q'_k} = \frac{(a_k + 1/a_{k+1})p'_{k-1} + p'_{k-2}}{(a_k + 1/a_{k+1})q'_{k-1} + q'_{k-2}} \\ &= \frac{(a_k + 1/a_{k+1})p_{k-1} + p_{k-2}}{(a_k + 1/a_{k+1})q_{k-1} + q_{k-2}} \\ &= \frac{a_{k+1}(a_k p_{k-1} + p_{k-2}) + p_{k-1}}{a_{k+1}(a_k q_{k-1} + q_{k-2}) + q_{k-1}} \\ &= \frac{a_{k+1}p_k + p_{k-1}}{a_{k+1}q_k + q_{k-1}} = \frac{p_{k+1}}{q_{k+1}}, \end{aligned}$$

as desired.  $\square$

**PROPOSITION 2.2.** *For any continued fraction  $[a_0; a_1, \dots]$  define the sequences  $p_k, q_k$  as in Proposition 2.1. Then for any  $k \geq 0$ ,  $p_k q_{k-1} - q_k p_{k-1} = (-1)^{k+1}$ .*

*Proof.* We will proceed by induction. For  $k = 0$ ,

$$p_0 q_{-1} - q_0 p_{-1} = a_0 \cdot 0 - 1 \cdot 1 = -1.$$

For  $k \geq 1$ , we have

$$\begin{aligned} p_k q_{k-1} - q_k p_{k-1} &= (a_k p_{k-1} + p_{k-2})q_{k-1} - (a_k q_{k-1} + q_{k-2})p_{k-1} \\ &= p_{k-2}q_{k-1} - q_{k-2}p_{k-1} \\ &= -(p_{k-1}q_{k-2} - q_{k-1}p_{k-2}) = -(-1)^k = (-1)^{k+1}. \end{aligned}$$

$\square$

## 3. ERGODIC THEORY

To study the statistical behavior of the Gauss map, we need a framework for analyzing the long-term behavior of dynamical systems. Ergodic theory provides exactly this. We assume familiarity with the Lebesgue measure and integral and will use them without further exposition. All integrals in this paper are Lebesgue integrals. For a more detailed introduction to Ergodic Theory, including proofs of the theorems, see [Sil08].

We begin by setting up the necessary measure-theoretic machinery. We start by defining the abstract framework of a measure space, which generalizes the notion of length to more complex sets.

**DEFINITION 3.1** ( $\sigma$ -algebra). Let  $X$  be a set, and let  $\Sigma$  be a set containing subsets of  $X$ . Then  $\Sigma$  is a  $\sigma$ -algebra if it satisfies the following properties:

- $\Sigma$  contains  $X$ .
- If some set  $A$  is in  $\Sigma$ , then so is  $X \setminus A$ .
- If a countable set of sets  $A_1, A_2, \dots$  is contained in  $\Sigma$ , then so is their union  $A_1 \cup A_2 \cup \dots$

**EXAMPLE 3.2.** For a given set  $X$ , let  $\Sigma$  be the set containing all subsets of  $X$ . Then  $\Sigma$  is called the power set of  $X$ , and it is a  $\sigma$ -algebra of  $X$ .

**DEFINITION 3.3** (Measure Space). Let  $X$  be a set and  $\Sigma$  be a  $\sigma$ -algebra over  $X$ . A function  $\mu : \Sigma \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$  is called a measure if it satisfies the following properties:

- $\mu(\emptyset) = 0$
- For all countable collections of pairwise disjoint sets  $\{A_i\}_{i=1}^{\infty}$ , we have

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

The triple  $(X, \Sigma, \mu)$  is called a measure space. If  $\mu(X) = 1$  then it is called a probability measure space.

**EXAMPLE 3.4.** Let  $X = \mathbb{R}$ ,  $\lambda$  be the 1-dimensional Lebesgue measure, and  $\Sigma$  consist of all Lebesgue measurable subsets of  $X$ . Then  $(X, \Sigma, \lambda)$  is a measure space.

**DEFINITION 3.5** (Almost Everywhere/Nowhere). For a measure space  $(X, \Sigma, \mu)$ , a property of  $X$  is said to hold almost nowhere if the set on which it holds has measure 0. If the set where it does not hold has measure 0, then the property is said to hold almost everywhere.

**EXAMPLE 3.6.** Consider the Lebesgue measure over  $\mathbb{R}$ . Almost every real number is irrational.

**DEFINITION 3.7** (Measurable function). Let  $(X, \Sigma, \mu)$  be a measure space, and let  $f : X \rightarrow \mathbb{R}$  be a function. If for all  $a \in \mathbb{R}$ , we have  $f^{-1}((a, \infty)) \in \Sigma$ , then we say that  $f$  is measurable.

**DEFINITION 3.8** ( $\mu$ -Integrable function). Let  $(X, \Sigma, \mu)$  be a measure space, and let  $f : X \rightarrow \mathbb{R}$  be a measurable function. If the Lebesgue integral  $\int_X |f| d\mu < \infty$ , then we say that  $f$  is  $\mu$ -integrable.

With a measure space in hand, we can now define transformations on it. The key properties we will care about are whether a transformation respects the measure structure.

**DEFINITION 3.9** (Measurable transformation). Let  $(X, \Sigma, \mu)$  be a measure space, and let  $T$  be a map from  $X$  to itself. We say that  $T$  is measurable if for any set  $A \in \Sigma$ ,  $T^{-1}(A) \in \Sigma$ .

**DEFINITION 3.10** (Measure-preserving transformation). Let  $(X, \Sigma, \mu)$  be a measure space, and let  $T$  be a measurable map from  $X$  to itself. We say that  $T$  is measure-preserving if for any set  $A \in \Sigma$ ,  $\mu(T^{-1}(A)) = \mu(A)$ .

**EXAMPLE 3.11.** Let  $X$  be the Euclidean plane and  $\mu$  be the 2-dimensional Lebesgue measure. Define  $T((x, y)) = (x + 1, y)$  to be translation to the right by 1 unit. Then  $T$  is measure-preserving as translating a set does not change its area.

Measure-preserving transformations can still behave in degenerate ways. For instance, they might leave some proper subregion entirely fixed. Ergodicity rules this out.

**DEFINITION 3.12** (Ergodic transformation). Let  $(X, \Sigma, \mu)$  be a measure space with  $\mu(X) < \infty$ , and let  $T$  be a measure-preserving transformation. We say that  $T$  is ergodic if for every  $A \in \Sigma$  satisfying  $T^{-1}(A) = A$ , either  $\mu(A) = 0$  or  $\mu(A) = \mu(X)$ .

One way to intuitively think of an ergodic transformation is like a spoon thoroughly stirring a glass of water. Ergodicity means that no subregion/droplet of water stays in place, and that the mixing is done in a way that does not change the density anywhere.

**EXAMPLE 3.13.** Let  $X = \mathbb{Z}/10\mathbb{Z}$ ,  $\Sigma$  be the power set of  $X$ , and  $\mu(A) = |A|$ . Then the transformation  $T(x) = x + 1$  is ergodic, as for any input  $x$ , the orbit  $\{x, T(x), T^2(x), \dots\}$  is all of  $X$ . By contrast, the map  $S(x) = x + 2$  is not ergodic as we have the  $S$ -invariant set  $\{0, 2, 4, 6, 8\}$  with measure 5.

We now present the key result from ergodic theory for this paper.

**THEOREM 3.14** (Birkhoff's Ergodic Theorem [Bir42]). *Let  $(X, \Sigma, \mu)$  be a measure space with  $\mu(X) < \infty$ , and let  $T$  be an ergodic transformation. Let  $f : X \rightarrow \mathbb{R}$  be a  $\mu$ -integrable function. Then for almost every  $x \in X$  we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \frac{1}{\mu(X)} \int_X f \, d\mu.$$

*The left hand side is commonly referred to as the time average, while the right hand side is the space average.*

#### 4. ERGODICITY OF THE GAUSS MAP

We now apply the framework of Section 3 to the Gauss map. The goal of this section is to show that the Gauss map is ergodic with respect to a natural probability measure on  $(0, 1)$ , called the Gauss measure, which will allow us to apply Birkhoff's Ergodic Theorem in the sections that follow.

The proofs in this section follow [Kar13, Chapter 9]. Let  $\bar{X} = (0, 1)$ , and  $\bar{\Sigma}$  be the Lebesgue measurable subsets of  $\bar{X}$ . From now on, we will work in the measure space  $X = \bar{X} \setminus \mathbb{Q}$ ,  $\Sigma = \{A \cap X \mid A \in \bar{\Sigma}\}$ , and define the Gauss measure  $\mu$  as

$$\mu(S) = \frac{1}{\ln 2} \int_S \frac{dx}{1+x}.$$

Note that excluding the rationals from our space does not affect any measures as  $\mathbb{Q}$  has Lebesgue measure 0. Also,  $\mu(X) = 1$ , so  $(X, \Sigma, \mu)$  is a probability measure space. From now on let  $T$  denote the Gauss map. Because  $X$  excludes all rationals, which are the only numbers with terminating continued fractions,  $T(X) = T^{-1}(X) = X$ .

We now proceed with the proof of the ergodicity of the Gauss Map.

PROPOSITION 4.1.  *$T$  is measure-preserving for the measure space  $(X, \Sigma, \mu)$ .*

*Proof.* For any  $S \in \Sigma$

$$\begin{aligned} \mu(T^{-1}(S)) &= \frac{1}{\ln 2} \int_{T^{-1}(S)} \frac{dx}{1+x} \\ &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \left( \int_{T^{-1}(S) \cap [1/(k+1), 1/k]} \frac{dx}{1+x} \right). \end{aligned}$$

For every  $k$ ,  $T$  creates a bijection from  $(\frac{1}{k+1}, \frac{1}{k})$  to  $(0, 1)$ , so we can define a single valued inverse  $T_k^{-1} : (0, 1) \rightarrow (1/(k+1), 1/k)$  of  $T$ . We have that  $T_k^{-1}(x) = \frac{1}{k+x}$ , and  $T^{-1}(S) \cap (1/(k+1), 1/k) = T_k^{-1}(S)$ .

Using this, for every integral in the sum, we can define  $u = T(x) = 1/x - k$ , which gives  $x = 1/(u+k)$  and  $dx = -du/(u+k)^2$ . Using this substitution gives

$$\int_{T^{-1}(S) \cap [1/(k+1), 1/k]} \frac{dx}{1+x} = \int_S \frac{du/(u+k)^2}{1+1/(u+k)} = \int_S \frac{du}{(u+k)(u+k+1)}.$$

Plugging this back into the sum gives

$$\begin{aligned} \mu(T^{-1}(S)) &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \left( \int_S \frac{dx}{(x+k)(x+k+1)} \right) \\ &= \frac{1}{\ln 2} \int_S \left( \sum_{k=1}^{\infty} \frac{1}{(x+k)(x+k+1)} \right) dx \quad (\text{the integrand is non-negative}) \\ &= \frac{1}{\ln 2} \int_S \left( \sum_{k=1}^{\infty} \left( \frac{1}{x+k} - \frac{1}{x+k+1} \right) \right) dx \\ &= \frac{1}{\ln 2} \int_S \left( \lim_{k \rightarrow \infty} \left( \frac{1}{x+1} - \frac{1}{x+k} \right) \right) dx \quad (\text{telescoping sum}) \\ &= \frac{1}{\ln 2} \int_S \frac{dx}{x+1} = \mu(S). \end{aligned}$$

We have showed that  $\mu(T^{-1}(S)) = \mu(S)$  for every measurable set  $S$ , so  $T$  is measure-preserving, as desired.  $\square$

We now present the most important property of the Gauss map.

THEOREM 4.2. *The Gauss map is Ergodic in  $(X, \Sigma, \mu)$ .*

The proof of this theorem is the most technical part of this section. The proof strategy is as follows: suppose  $S$  is a  $T$ -invariant set of positive measure. We want to show  $\mu(S) = 1$ . To do this, we show that  $\mu(X \setminus S) = 0$ , by arguing via Lebesgue's Density Theorem that  $X \setminus S$  is too sparse to have positive measure. We first state the density theorem, then establish two lemmas controlling the size of  $S$  inside fundamental intervals, before assembling the full proof.

**THEOREM 4.3** (Lebesgue's Density Theorem). *Let  $\lambda$  denote the 1-dimensional Lebesgue measure and let  $B(x, \epsilon)$  be the 1-dimensional ball with center  $x$  and radius  $\epsilon$ . Let  $A$  be a set in  $\Sigma$ . For any  $x \in A$ , define the Lebesgue density of  $x$  as*

$$\lim_{\epsilon \rightarrow 0} \frac{\lambda(A \cap B(x, \epsilon))}{\lambda(B(x, \epsilon))}.$$

*Then we have that almost every  $x \in A$  has Lebesgue density 1.*

A detailed proof of this theorem can be found in [Mat95, Corollary 2.14]. Note that the measures  $\lambda$  and  $\mu$  are equivalent, i.e.  $\mu(A) = 0 \iff \lambda(A) = 0$ . This gives us the following corollary.

**COROLLARY 4.4.** *Let  $A$  be a set in  $\Sigma$ . For  $x \in A$ , define the  $\mu$ -density of  $x$  as*

$$\lim_{\epsilon \rightarrow 0} \frac{\mu(A \cap B(x, \epsilon))}{\mu(B(x, \epsilon))}.$$

*Then almost every  $x \in A$  has  $\mu$ -density 1.*

*Proof Sketch.* Since  $\mu$  and  $\lambda$  are equivalent and  $\mu$  has a continuous, strictly positive density, the measures are locally proportional as  $\epsilon \rightarrow 0$ . This local density factor cancels out in the ratio, effectively recovering the classical Lebesgue result.  $\square$

We will now show the lemmas that provide the local estimates needed to apply Corollary 4.4. First, we will need to introduce additional notation.

For a tuple of positive integers  $(a_1, \dots, a_n)$ , define the set  $I_{(a_1, \dots, a_n)}$  as the open interval with endpoints  $[a_1, \dots, a_n]$  and  $[a_1, \dots, a_n + 1]$ . From the properties of the Gauss map, we know that the restriction of the map

$$T^n : I_{(a_1, \dots, a_n)} \rightarrow (0, 1)$$

is bijective. Define its inverse

$$F_{(a_1, \dots, a_n)} : (0, 1) \rightarrow I_{(a_1, \dots, a_n)},$$

$$F_{(a_1, \dots, a_n)}(x) = [a_1, \dots, a_n, 1/x].$$

If  $p_k/q_k$  is the  $k$ th convergent of  $[a_1, \dots, a_n]$ , then from Proposition 2.1

$$F_{(a_1, \dots, a_n)}(x) = \frac{p_n/x + p_{n-1}}{q_n/x + q_{n-1}} = \frac{p_n + p_{n-1}x}{q_n + q_{n-1}x}.$$

**LEMMA 4.5.** *For any  $n$ -tuple  $(a_1, \dots, a_n)$  of positive integers, with even  $n$ , we have*

$$\mu(I_{(a_1, \dots, a_n)}) < \frac{1}{\ln 2(q_n + q_{n-1})(p_n + q_n)}.$$

*Proof.* We have

$$\begin{aligned}
\mu(I_{(a_1, \dots, a_n)}) &= \frac{1}{\ln 2} \int_{I_{(a_1, \dots, a_n)}} \frac{dx}{1+x} \\
&= \frac{1}{\ln 2} \left| \int_{[a_1, \dots, a_{n-1}]}^{[a_1, \dots, a_n]} \frac{dx}{1+x} \right| \\
&= \frac{1}{\ln 2} \left| \ln \left( \left(1 + \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right) / \left(1 + \frac{p_n}{q_n}\right) \right) \right| && \text{(by Proposition 2.1)} \\
&= \frac{1}{\ln 2} \left| \ln \left( 1 + \frac{q_n p_{n-1} - p_n q_{n-1}}{(q_n + q_{n-1})(p_n + q_n)} \right) \right| \\
&= \frac{1}{\ln 2} \left| \ln \left( 1 + \frac{(-1)^n}{(q_n + q_{n-1})(p_n + q_n)} \right) \right| && \text{(by Proposition 2.2)} \\
&= \frac{1}{\ln 2} \left| \ln \left( 1 + \frac{1}{(q_n + q_{n-1})(p_n + q_n)} \right) \right| && \text{(as } n \text{ is even)} \\
&< \frac{1}{\ln 2 (q_n + q_{n-1})(p_n + q_n)} && \text{(using } \ln(1+x) < x \text{).}
\end{aligned}$$

□

The following lemma provides the local estimate needed to conclude sparsity of  $X \setminus S$  and apply Corollary 4.4.

LEMMA 4.6. *Let  $S$  be a set with positive measure that is invariant under  $T$ . For any interval  $I_{(a_1, \dots, a_n)}$ , with even  $n$ ,*

$$\mu(S \cap I_{(a_1, \dots, a_n)}) \geq \frac{\ln(2)}{2} \mu(S) \mu(I_{(a_1, \dots, a_n)}).$$

*Proof.* We know that  $T^{-1}(S) = S$ , and that  $T$  is surjective onto  $[0, 1)$ , so  $T(S) = S$ . Hence,  $T^n(S) = S$ . From the properties of the Gauss Map, we know that  $T^n$  is bijective when restricted to  $I_{(a_1, \dots, a_n)}$ , so  $T^n(S \cap I_{(a_1, \dots, a_n)}) = S$ . For any  $x \in I_{(a_1, \dots, a_n)}$ , let  $u = T^n(x)$ , which gives

$$x = F_{(a_1, \dots, a_n)}(u) = \frac{p_n + p_{n-1}u}{q_n + q_{n-1}u}, \text{ and}$$

$$dx = \frac{(q_n + q_{n-1}u)p_{n-1} - (p_n + p_{n-1}u)q_{n-1}}{(q_n + q_{n-1}u)^2} du = \frac{(-1)^n}{(q_n + q_{n-1}u)^2} du.$$

Using this substitution, we have

$$\begin{aligned}
\frac{1}{\ln 2} \int_{S \cap I_{(a_1, \dots, a_n)}} \frac{dx}{1+x} &= \frac{1}{\ln 2} \left| \int_S \frac{(-1)^n}{(q_n + q_{n-1}u)^2} \frac{du}{1 + (p_n + p_{n-1}u)/(q_n + q_{n-1}u)} \right| \\
&= \frac{1}{\ln 2} \int_S \frac{du}{(q_n + q_{n-1}u)(q_n + q_{n-1}u + p_n + p_{n-1}u)}
\end{aligned}$$

since  $p_n > p_{n-1}$  and  $q_n > q_{n-1}$  by Proposition 2.1,

$$\begin{aligned}
 &\geq \frac{1}{q_n(p_n + q_n)} \frac{1}{\ln 2} \int_S \frac{du}{(1+u)^2} \\
 &\geq \frac{1}{q_n(p_n + q_n)} \frac{1}{2 \ln 2} \int_S \frac{du}{1+u} \\
 &= \frac{1}{2q_n(p_n + q_n)} \mu(S) \\
 &\geq \frac{1}{2(q_n + q_{n-1})(p_n + q_n)} \mu(S) \\
 &> \frac{\ln(2)}{2} \mu(S) \mu(I_{(a_1, \dots, a_n)}),
 \end{aligned}$$

where we get the last inequality from Lemma 4.5.  $\square$

We now assemble the proof of Theorem 4.2.

*Proof of Theorem 4.2.* Let  $\lambda$  denote the 1-dimensional Lebesgue measure. Let  $S$  be a set satisfying  $T^{-1}(S) = S$  with positive measure. For any irrational  $x = [a_1, a_2, \dots]$  and even  $n$ ,

$$x \in I_{(a_1, a_2, \dots, a_n)} \implies I_{(a_1, a_2, \dots, a_n)} \subseteq B(x, \lambda(I_{(a_1, a_2, \dots, a_n)})).$$

We know that for  $x \in (0, 1)$ ,  $\frac{1}{2} \leq \frac{1}{1+x} \leq 1$ , which implies that for any  $A \in \Sigma$ ,  $\frac{1}{2 \ln(2)} \lambda(S) \leq \mu(S) \leq \frac{1}{\ln(2)} \lambda(S)$ . Using these bounds, along with Lemma 4.6, we get

$$\begin{aligned}
 \frac{\mu((X \setminus S) \cap B(x, \lambda(I_{(a_1, \dots, a_n)})))}{\mu(B(x, \lambda(I_{(a_1, \dots, a_n)})))} &= 1 - \frac{\mu(S \cap B(x, \lambda(I_{(a_1, \dots, a_n)})))}{\mu(B(x, \lambda(I_{(a_1, \dots, a_n)})))} \\
 &\leq 1 - \frac{\ln(2)}{2} \frac{\mu(S) \mu(I_{(a_1, \dots, a_n)})}{\mu(B(x, \lambda(I_{(a_1, \dots, a_n)})))} \\
 &\leq 1 - \frac{\ln(2)}{2} \frac{\mu(S) \lambda(I_{(a_1, \dots, a_n)}) / 2 \ln(2)}{\lambda(B(x, \lambda(I_{(a_1, \dots, a_n)}))) / \ln(2)} \\
 &= 1 - \frac{\ln(2)}{4} \frac{\mu(S) \lambda(I_{(a_1, \dots, a_n)})}{\lambda(B(x, \lambda(I_{(a_1, \dots, a_n)})))} \\
 &= 1 - \frac{\ln(2)}{8} \mu(S) < 1.
 \end{aligned}$$

By taking  $n \rightarrow \infty$ , the lengths of the intervals  $I_{(a_1, \dots, a_n)}$  go to 0, so we have that every irrational number has  $\mu$ -density less than 1 in  $X \setminus S$ . Because  $X = (0, 1) \setminus \mathbb{Q}$ , we know  $X \setminus S$  contains no rationals, so no points in  $X \setminus S$  have  $\mu$ -density 1. However, Corollary 4.4 tells us that almost every point in  $X \setminus S$  has  $\mu$ -density 1. Thus,  $\mu(X \setminus S) = 0 \implies \mu(S) = 1$ , implying that  $T$  is ergodic.  $\square$

## 5. THE GAUSS-KUZMIN THEOREM

We now arrive at the central result of this paper: the Gauss–Kuzmin distribution, which describes the limiting frequency of digits in the continued fraction expansion of a typical real number.

**5.1. Pointwise Gauss-Kuzmin Theorem.** We begin with a pointwise version of the Gauss–Kuzmin theorem, which follows naturally from the ergodicity of the Gauss map.

Let  $x \in (0, 1)$  have continued fraction expansion  $[a_1, a_2, \dots]$ . For positive integers  $n$  and  $k$ , we define

$$P_{n,k}(x) = \frac{\#\{i \leq n \mid a_i = k\}}{n},$$

the proportion of the first  $n$  digits of  $x$  equal to  $k$ , and (when the limit exists)

$$P_k(x) = \lim_{n \rightarrow \infty} P_{n,k}(x).$$

Thus,  $P_k(x)$  represents the asymptotic frequency of the digit  $k$  along the single orbit of  $x$  under the Gauss map.

**THEOREM 5.1** (Pointwise Gauss-Kuzmin Theorem [Kar13, Theorem 9.14]). *For almost every  $x \in (0, 1)$ , the limit  $P_k(x)$  exists for all  $k \geq 1$ , and is given by*

$$P_k(x) = \frac{1}{\ln 2} \ln \left( 1 + \frac{1}{k(k+2)} \right).$$

*Proof.* For any positive integer  $k$ , define

$$f_k(x) = \begin{cases} 1, & \text{if } x \in (\frac{1}{k+1}, \frac{1}{k}) \\ 0, & \text{otherwise.} \end{cases}$$

Note that for any  $x \in (0, 1)$  with continued fraction expansion  $[a_1, a_2, \dots]$ , we have  $a_1 = k \iff f_k(x) = 1$ . Thus,

$$P_{n,k}(x) = \frac{1}{n} \sum_{i=0}^{n-1} f_k(T^i(x)).$$

We know that  $T$  is ergodic (Theorem 4.2). Using Birkhoff’s Ergodic Theorem (Theorem 3.14), we have that for almost every  $x \in X$

$$\begin{aligned} P_k(x) &= \lim_{n \rightarrow \infty} P_{n,k}(x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f_k(T^i(x)) \\ &= \frac{1}{\mu(X)} \int_X f_k d\mu \\ &= \frac{1}{\ln 2} \int_{1/(k+1)}^{1/k} \frac{dx}{1+x} \\ &= \frac{1}{\ln 2} \left[ \ln \left( 1 + \frac{1}{k} \right) - \ln \left( 1 + \frac{1}{k+1} \right) \right] \\ &= \frac{1}{\ln 2} \ln \left( 1 + \frac{1}{k(k+2)} \right). \end{aligned}$$

We are done as  $X$  contains almost every point in  $(0, 1)$ . □

**5.2. Original Gauss-Kuzmin Theorem.** The pointwise theorem tells us that digit frequencies converge for almost every  $x$ , but says nothing about how quickly. The classical Gauss-Kuzmin theorem, due to Wirsing, strengthens this by giving a quantitative exponential rate of convergence for the distribution of  $T^n(x)$  across all of  $(0, 1)$ , not just for a single typical point. This is a significantly deeper result, and its proof requires tools from operator theory that lie beyond the scope of this paper. We state it here for completeness and because it implies a clean corollary about digit frequencies.

**THEOREM 5.2** (Gauss-Kuzmin Theorem [Wir74]). *For all  $x \in (0, 1), n \in \mathbb{N}$ , define*

$$m_n(x) = \lambda(\{y \in (0, 1) \setminus \mathbb{Q} \mid T^n(y) < x\}),$$

where  $\lambda$  is the 1-dimensional Lebesgue measure. Then there exists a real constant  $c > 0$  such that for all  $x \in (0, 1), n \in \mathbb{N}$

$$|m_n(x) - \log_2(1+x)| \leq c(0.31)^n.$$

In other words,  $m_n(x)$  exponentially converges to  $\log_2(1+x)$ .

This gives the following corollary.

**COROLLARY 5.3.** *For all positive integers  $n, k$ , define*

$$P_n(k) = \lambda(\{x \in (0, 1) \setminus \mathbb{Q} \mid a_n(x) = k\}),$$

where  $a_n(x)$  is the  $n$ th term in the continued fraction representation of  $x$ , and  $\lambda$  is the 1-dimensional Lebesgue measure. Let

$$P(k) = \lim_{n \rightarrow \infty} P_n(k).$$

Then  $P(k) = \log_2\left(1 + \frac{1}{k(k+2)}\right)$ .

*Proof.* Note that  $P_n(k) = m_{n-1}(1/k) - m_{n-1}(1/(k+1))$ , so

$$\begin{aligned} P(k) &= \lim_{n \rightarrow \infty} m_{n-1}(1/k) - m_{n-1}(1/(k+1)) \\ &= \log_2(1 + 1/k) - \log_2(1 + 1/(k+1)) = \log_2\left(1 + \frac{1}{k(k+2)}\right). \end{aligned}$$

□

## 6. KHINCHIN'S CONSTANT

The Gauss-Kuzmin theorem characterizes how often each digit appears in a typical continued fraction expansion. A natural follow-up question is: what can we say about the size of the digits? It turns out that the geometric mean of the continued fraction digits converges to the same constant for almost every real number, another clean consequence of the ergodicity of the Gauss map.

**THEOREM 6.1** ([RN51]). *For almost every real number  $x$ ,*

$$\lim_{n \rightarrow \infty} (a_1 a_2 \cdots a_n)^{1/n} = K_0,$$

where  $[a_0; a_1, \dots]$  is the continued fraction expansion of  $x$  and  $K_0 \approx 2.685$  is known as Khinchin's constant.

*Proof.* The term  $a_0$  has no importance, and the rationals have measure 0, so we can focus on the same space  $X = (0, 1) \setminus \mathbb{Q}$  as before. Define  $f : X \rightarrow \mathbb{R}$  with  $f(x) = \ln[1/x]$ . Note that  $f([a_1, a_2, \dots]) = \ln a_1$ . As  $T$  is ergodic, Birkhoff's Ergodic Theorem tells us that for almost every  $x \in X$  (and as a consequence for almost every  $x \in \mathbb{R}$ ), we have

$$\begin{aligned}
\ln \left( \lim_{n \rightarrow \infty} (a_1 a_2 \cdots a_n)^{1/n} \right) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \ln(a_i) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) && \text{(using } f([a_1, a_2, \dots]) = \ln a_1 \text{)} \\
&= \frac{1}{\mu(X)} \int_X f d\mu && \text{(by Theorem 3.14)} \\
&= \sum_{k=1}^{\infty} \int_{1/(k+1)}^{1/k} f d\mu && \text{(the intervals partition } X \text{)} \\
&= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \int_{1/(k+1)}^{1/k} \frac{f(x) dx}{1+x} && \text{(by the definition of } \mu \text{)} \\
&= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \int_{1/(k+1)}^{1/k} \frac{\ln(k) dx}{1+x} && \text{(using } f([a_1, a_2, \dots]) = \ln a_1 \text{)} \\
&= \sum_{k=1}^{\infty} \frac{\ln(k)}{\ln 2} \ln \left( 1 + \frac{1}{k(k+2)} \right) \\
&= \ln \left( \prod_{k=1}^{\infty} \left( 1 + \frac{1}{k(k+2)} \right)^{\log_2(k)} \right) \\
&= \ln(K_0) && \text{(definition of } K_0 \text{).}
\end{aligned}$$

Exponentiating both sides completes the proof.  $\square$

#### REFERENCES

- [Bir42] G. D. Birkhoff. What is the ergodic theorem? *Amer. Math. Monthly*, 49:222–226, 1942.
- [Kar13] Oleg Karpenkov. *Geometry of continued fractions*, volume 26 of *Algorithms and Computation in Mathematics*. Springer, Heidelberg, 2013.
- [Mat95] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.
- [RN51] C. Ryll-Nardzewski. On the ergodic theorems. II. Ergodic theory of continued fractions. *Studia Math.*, 12:74–79, 1951.
- [Sil08] C. E. Silva. *Invitation to ergodic theory*, volume 42 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2008.
- [Wir74] Eduard Wirsing. On the theorem of Gauss-Kusmin-Lévy and a Frobenius-type theorem for function spaces. *Acta Arith.*, 24:507–528, 1973/74.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
Email address: paulmte5@mit.edu