THE GAUSS MAP AND THE DISTRIBUTION OF PARTIAL QUOTIENTS

FIONA LEE

Abstract. Every irrational number has a unique continued fraction expansion, and the partial quotients of this expansion govern how well the number can be approximated by rationals. This expository paper gives a self-contained account of the statistical behavior of partial quotients for a typical real number. We introduce the Gauss map $T(x) = \{1/x\}$ (where $\{\cdot\}$ denotes the fractional part), which repackages the continued fraction algorithm as iteration of a single function on $(0, 1]$. We prove that T preserves the Gauss measure $d\mu = \frac{1}{\ln 2} \cdot \frac{dx}{1+x}$, sketch the proof of the Gauss–Kuzmin–Lévy theorem on the convergence of the distribution of remainders, and derive Khinchin's constant from it. The exposition introduces the necessary measure-theoretic background from scratch; the deeper ergodic-theoretic input needed for the almost-everywhere statements is identified but not proved.

Contents

1. Introduction

The continued fraction expansion of an irrational number $\alpha \in (0, 1)$ is the expression

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} = [0; a_1, a_2, a_3, \ldots],$$

where $a_1, a_2, a_3, \ldots$ are positive integers called the *partial quotients* of α. The rational approximations obtained by truncating this expansion, the *convergents* $p_k/q_k = [0; a_1, \ldots, a_k]$, satisfy the recurrences

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

with $p_0 = 0$, $q_0 = 1$, $p_1 = 1$, $q_1 = a_1$. The convergents p_k/q_k are the best rational approximations to α in the following sense: p_k/q_k is the closest fraction to α among

Date: MIT 18.784, Spring 2026.

all fractions whose denominator is at most q_k. The quality of the kth convergent is controlled by the next partial quotient a_{k+1}: the standard error estimate gives $|\alpha - p_k/q_k| \approx 1/(a_{k+1}q_k^2)$, so a large partial quotient a_{k+1} acts as a multiplier on the quality of the approximation p_k/q_k.

A natural question arises: what do the partial quotients of a typical real number look like? Are they mostly small, or do large values appear frequently? There are two ways to approach this question, and they give different kinds of answers. Classical Diophantine approximation asks what is true for *every* irrational. For instance, Dirichlet's theorem [11] guarantees infinitely many rational approximations p/q with $|\alpha - p/q| < 1/q^2$ for any irrational α, and Hurwitz [12] sharpened the constant to $1/(\sqrt{5}\,q^2)$. The Dirichlet and Hurwitz bounds are universal but coarse. The finer statistical behavior of the partial quotients requires a different perspective, one that asks what happens for *almost every* real number. This is a measure-theoretic notion, not a colloquial one—it means the set of exceptions has Lebesgue measure zero, i.e., can be covered by intervals of arbitrarily small total length (we will develop this formally in Section 3). The metrical results are much sharper but come with a caveat: they say nothing about any particular α.

The central result of this paper is the **Gauss–Kuzmin theorem**: for almost every irrational α, the proportion of partial quotients equal to k converges to

$$d(k) = \log_2\left(1 + \frac{1}{k(k+2)}\right). \tag{1.1}$$

This means that roughly 41.5% of partial quotients are 1, about 17% are 2, about 9.3% are 3, and so on. Strikingly, no specific constant of mathematical interest (such as π or $\log 2$) has been proved to obey this law, though computational experiments with up to 10^8 partial quotients of π show excellent agreement [2, Section 3.4], as illustrated in Table 1. For algebraic irrationals of degree ≥ 3 the question is wide open, and recent work suggests the law may not apply uniformly to them [10].

k	1	2	3	4	5	6	7	8	9	10
$d(k)$	.415	.170	.093	.059	.041	.030	.023	.018	.014	.012
π	.415	.170	.093	.059	.041	.030	.023	.018	.015	.012

Table 1. Theoretical frequencies $d(k)$ vs. observed frequencies from the first 10^8 partial quotients of π [2, Section 3.4].

The Gauss–Kuzmin distribution also determines the long-run behavior of products and other averages of partial quotients. The most striking consequence is the existence of *Khinchin's constant*: for almost every α, the geometric mean of the first N partial quotients converges to a universal constant,

$$\lim_{N\to\infty} (a_1 a_2 \cdots a_N)^{1/N} = K_0 = 2.685\,452\ldots,$$

regardless of α. Meanwhile, the partial quotients themselves grow without bound—for almost every α, a_n exceeds any fixed threshold infinitely often.

The key to proving these results is the **Gauss map** $T(x) = \{1/x\}$, which repackages the continued fraction algorithm as iteration of a single function on $(0, 1]$. Iterating T generates the partial quotients one at a time, so questions about their long-run statistics become questions about the dynamics of T. The crucial fact is that T preserves a natural probability measure, the **Gauss measure** $d\mu = \frac{1}{\ln 2} \cdot \frac{dx}{1+x}$. Because μ is invariant, the distribution of the remainders $T^n(\alpha)$ converges to μ regardless of the starting distribution, and the Gauss–Kuzmin frequencies follow. One further ingredient, a mixing property of T called *ergodicity* (explained in Section 4), upgrades these distributional statements to almost-everywhere results by forcing the time average along any single orbit to match the global average. Section 2 introduces the Gauss map, Section 3 develops the necessary measure theory, Section 4 proves the Gauss–Kuzmin theorem, and Section 5 derives Khinchin's constant.

2. The Gauss map as the continued fraction algorithm

The continued fraction algorithm is usually described as a sequence of steps: take the reciprocal, extract the integer part, repeat. The goal of this section is to repackage all of these steps into iteration of a single function. This point of view transforms questions about the statistics of partial quotients into questions about the long-term behavior of a dynamical system, which we can then attack with measure-theoretic tools.

Definition 2.1. The *Gauss map* is the function $T : (0, 1] \to [0, 1)$ defined by

$$T(x) = \left\{ \frac{1}{x} \right\} = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor .$$

The point of the Gauss map is that iterating T recovers the continued fraction expansion of any irrational $\alpha \in (0, 1)$. To see this, define the sequence of *remainders* $x_0, x_1, x_2, \dots$ by $x_0 = \alpha$ and $x_{n+1} = T(x_n)$ for $n \geq 0$, and define the *partial quotients* by

$$a_{n+1} = \left\lfloor \frac{1}{x_n} \right\rfloor .$$

Since $T(x_n) = 1/x_n - a_{n+1}$, we can write $x_n = 1/(a_{n+1} + x_{n+1})$. Unrolling this recurrence gives

$$\alpha = x_0 = \frac{1}{a_1 + x_1} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + x_2}} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}},$$

which is exactly the continued fraction expansion of α. In short: each application of T extracts one partial quotient and passes the remainder to the next step.

Example 2.2. To see the Gauss map in action, let $\alpha = \sqrt{2} - 1$ and set $x_0 = \alpha$. Then

$$\frac{1}{x_0} = \frac{1}{\sqrt{2} - 1} = \sqrt{2}+1, \qquad a_1 = \lfloor \sqrt{2}+1 \rfloor = 2, \qquad x_1 = (\sqrt{2}+1)-2 = \sqrt{2}-1 = x_0.$$

In terms of the Gauss map, $T(x_0) = (\sqrt{2} + 1) - 2 = \sqrt{2} - 1 = x_0$, so the orbit $x_0, T(x_0), T^2(x_0), \dots$ is the constant sequence $\sqrt{2}-1, \sqrt{2}-1, \dots$, and each iteration extracts the same partial quotient $a_n = 2$. This gives $\sqrt{2} - 1 = [0; 2, 2, 2, \dots]$, or equivalently $\sqrt{2} = [1; 2, 2, 2, \dots]$.

Numbers like $\sqrt{2}$ and e have predictable partial quotients ($\sqrt{2}$ repeats the same value forever, e follows a rigid pattern due to Euler). But most real numbers are not this well-behaved, and we cannot hope to describe their partial quotients individually. Instead, we can ask a statistical question. If we pick a real number "at random," what does the typical sequence of partial quotients look like? The graph of T (Figure 1) gives a first hint.

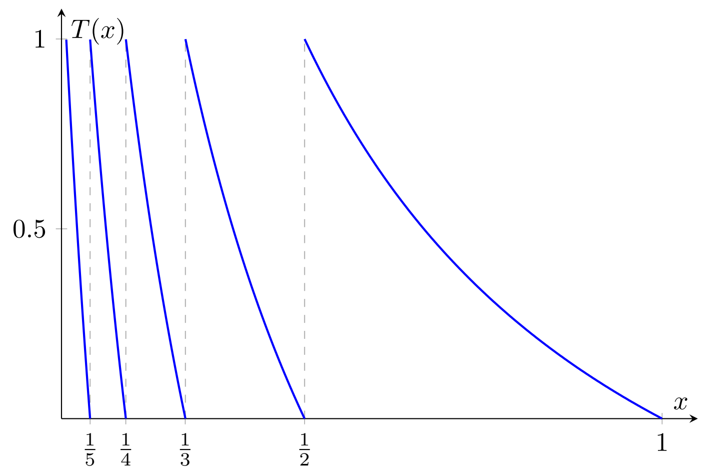

Figure 1. The graph of the Gauss map $T(x) = \{1/x\}$ on $(0, 1]$. On each interval $(1/(k+1), 1/k]$, the map takes the form $T(x) = 1/x - k$, which is a decreasing branch from 1 down to 0.

On the interval $\left(\frac{1}{k+1}, \frac{1}{k}\right]$, the partial quotient a_1 is constant equal to k, and the interval $\left(\frac{1}{k+1}, \frac{1}{k}\right]$ has length $\frac{1}{k(k+1)}$. So if we pick α uniformly at random from $(0, 1)$, then α has first partial quotient k with probability $\frac{1}{k(k+1)}$. But the uniform distribution only tells us about the first partial quotient; the distribution of the second, third, and subsequent partial quotients turns out to be more delicate. To understand the long-run frequencies, we will need a probability measure on $(0, 1)$ that T does not distort—an *invariant* measure.

3. Measures on $(0, 1)$

The results of this paper are statements about "almost every" real number, and their proofs rest on finding the right way to measure subsets of $(0, 1)$. This section develops the two measures we need. First, we introduce Lebesgue measure, which formalizes the notion of length and makes "almost every" precise. We then define fundamental intervals, which connect continued fractions to measure theory by partitioning $(0, 1)$ according to the first n partial quotients. Finally, we introduce the Gauss measure, the invariant measure for the Gauss map.

3.1. Lebesgue measure.

For a thorough treatment of the concepts in this subsection, see [4, §1.2].

Definition 3.1. The *Lebesgue measure* of an interval $(a, b) \subseteq (0, 1)$ is its length: $\mathcal{M}\big((a, b)\big) = b - a$. More generally, if S is a subset of $(0, 1)$ that can be covered by countably many intervals $I_1, I_2, \ldots$, then

$$\mathcal{M}(S) = \inf\left\{\sum_{n=1}^{\infty} |I_n| : S \subseteq \bigcup_{n=1}^{\infty} I_n\right\},$$

where $|I_n|$ denotes the length of I_n.

Not every subset of $(0,1)$ is Lebesgue measurable, but all sets built from countable unions, intersections, and complements of intervals are, and these are the only sets we will encounter.

Two key properties of Lebesgue measure that we will use throughout:

(i) **Countable additivity.** If $S_1, S_2, \ldots$ are pairwise disjoint measurable sets, then $\mathcal{M}\big(\bigcup S_n\big) = \sum \mathcal{M}(S_n)$.

(ii) **Countable subadditivity.** For any measurable sets $S_1, S_2, \ldots$ (not necessarily disjoint), $\mathcal{M}\big(\bigcup S_n\big) \le \sum \mathcal{M}(S_n)$.

Definition 3.2. A set $S \subseteq (0,1)$ has *measure zero* if for every $\varepsilon > 0$, there exist countably many intervals $I_1, I_2, \ldots$ with $S \subseteq \bigcup I_n$ and $\sum |I_n| < \varepsilon$. Equivalently, $\mathcal{M}(S) = 0$.

Any countable set has measure zero (cover the nth point by an interval of length $\varepsilon/2^n$). For example, the rationals in $(0,1)$ form a set of measure zero. When we say a property holds for *almost every* $\alpha \in (0,1)$, we mean the set of exceptions has measure zero.

Why is this the right notion of "typical"? Weaker alternatives like "all but countably many" are insufficient: the exceptional sets in continued fraction theory are typically uncountable yet negligibly small. For example, the set of α whose partial quotients are all at most 2 is uncountable (it contains $[0; d_1, d_2, \ldots]$ for every sequence $d_i \in \{1,2\}$), yet has Lebesgue measure zero.

3.2. **Fundamental intervals.** The connection between continued fractions and measure theory runs through the following construction. Given positive integers $k_1, \ldots, k_n$, define the *fundamental interval* (or *interval of rank n*) to be

$$I(k_1, \ldots, k_n) = \{\alpha \in (0,1) : a_1(\alpha) = k_1, \ldots, a_n(\alpha) = k_n\}.$$

This is the set of all irrationals in $(0,1)$ whose continued fraction expansion begins with $k_1, \ldots, k_n$.

Lemma 3.3 ([1, Chapter III, §12]). *$I(k_1, \ldots, k_n)$ is an interval with endpoints*

$$\frac{p_n}{q_n} \quad \text{and} \quad \frac{p_n + p_{n-1}}{q_n + q_{n-1}},$$

where $p_i/q_i = [0; k_1, \ldots, k_i]$ are the convergents. Its length is

$$|I(k_1, \ldots, k_n)| = \frac{1}{q_n(q_n + q_{n-1})}. \tag{3.1}$$

Idea. Any $\alpha \in I(k_1, \ldots, k_n)$ can be written as $\alpha = [0; k_1, \ldots, k_n, r_{n+1}]$ for some $r_{n+1} \in [1, \infty)$. A standard identity from the theory of continued fractions expresses this as

$$\alpha = \frac{p_n\, r_{n+1} + p_{n-1}}{q_n\, r_{n+1} + q_{n-1}},$$

which is a strictly monotonic function of r_{n+1} (the derivative is $(-1)^n/(q_n r_{n+1} + q_{n-1})^2$, nonzero by the convergent identity $p_n q_{n-1} - p_{n-1} q_n = (-1)^n$). Letting r_{n+1}

range over $[1, \infty)$ traces out an interval whose endpoints are obtained at $r_{n+1} = 1$ and $r_{n+1} \to \infty$, giving $(p_n + p_{n-1})/(q_n + q_{n-1})$ and p_n/q_n respectively. Subtracting the endpoints and using the same convergent identity gives the length formula. $\square$

As $(k_1, \ldots, k_n)$ ranges over all n-tuples of positive integers, the fundamental intervals $I(k_1, \ldots, k_n)$ are pairwise disjoint and their union covers every irrational in $(0, 1)$. The only points missed are rationals whose continued fraction expansion has length at most n, and there are only countably many such points, so the missed set has measure zero. In short: the rank-n fundamental intervals partition $(0, 1)$ up to a measure-zero set.

The length formula (3.1) already tells us how much of $(0, 1)$ is occupied by numbers with a given prefix. But for the arguments in later sections, what matters most is not the absolute length of a fundamental interval but the relative proportion: if we know that α lies in $I(k_1, \ldots, k_n)$, what fraction of that interval corresponds to having $a_{n+1} = k$? This is measured by the ratio $|I(k_1, \ldots, k_n, k)|/|I(k_1, \ldots, k_n)|$, and a key estimate (which follows from (3.1) and the recurrence for q_n; see [1, Chapter III, §12]) pins it down:

$$\frac{1}{3k^2} < \frac{|I(k_1, \ldots, k_n, k)|}{|I(k_1, \ldots, k_n)|} < \frac{2}{k^2}. \tag{3.2}$$

In words, the conditional probability that $a_{n+1} = k$, given any values of $a_1, \ldots, a_n$, is always on the order of $1/k^2$. This uniformity is the intuitive reason a universal frequency law exists: no matter what partial quotients came before, the next one is drawn from roughly the same distribution.

3.3. **The Gauss measure.** Lebesgue measure lets us formalize "almost every," but applying T distorts the uniform distribution, so it is not the right measure for studying the dynamics of T. We need a probability measure that is *invariant* under T, meaning T does not change it. Gauss identified such a measure in a letter to Laplace [9].

Definition 3.4. The *Gauss measure* on $(0, 1)$ is the probability measure defined by

$$\mu(A) = \frac{1}{\ln 2} \int_A \frac{dx}{1 + x}$$

for any measurable set $A \subseteq (0, 1)$.

The normalizing constant $1/\ln 2$ ensures that $\mu\big((0, 1)\big) = \frac{1}{\ln 2} \int_0^1 \frac{dx}{1+x} = \frac{\ln 2}{\ln 2} = 1$. Its distribution function is $\mu([0, x]) = \frac{1}{\ln 2} \ln(1 + x) = \log_2(1 + x)$. The Gauss measure assigns slightly more weight to the left end of $(0, 1)$ than to the right end (since $1/(1 + x)$ is decreasing), reflecting the fact that the Gauss map spends more time near 0.

We now verify that μ is indeed invariant under T. The strategy is to show that for any interval $[0, x]$, the Gauss measure of its preimage $T^{-1}([0, x])$ equals $\mu([0, x])$. Since T acts by a different formula on each interval $(1/(k + 1), 1/k]$, the preimage $T^{-1}([0, x])$ is a union of infinitely many subintervals (one for each $k \geq 1$), and computing its measure reduces to a sum over k that telescopes.

Theorem 3.5 (Invariance of the Gauss measure [1, Chapter III, §15]). *The Gauss map preserves the Gauss measure: for any measurable set $A \subseteq (0, 1)$,*

$$\mu(T^{-1}(A)) = \mu(A).$$

Proof. It suffices to check this for intervals $A = [0, x]$ with $0 \le x \le 1$, since every measurable subset of $(0, 1)$ can be built from such intervals by countable unions, intersections, and complements, and both sides of the equation respect these operations. We compute $T^{-1}([0, x])$ explicitly. For each positive integer k, the condition $T(\alpha) \le x$ with $a_1(\alpha) = k$ means $1/\alpha - k \le x$, i.e., $\alpha \ge 1/(k + x)$. Combined with the constraint $\alpha \in \left(\frac{1}{k+1}, \frac{1}{k}\right]$, this gives

$$T^{-1}([0, x]) \cap \left(\frac{1}{k+1}, \frac{1}{k}\right] = \left[\frac{1}{k+x}, \frac{1}{k}\right].$$

Therefore

$$\begin{aligned} \mu(T^{-1}([0, x])) &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \int_{1/(k+x)}^{1/k} \frac{dt}{1+t} \\ &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \left[\ln\left(1 + \frac{1}{k}\right) - \ln\left(1 + \frac{1}{k+x}\right)\right] \\ &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \left[\ln\left(\frac{k+1}{k}\right) - \ln\left(\frac{k+1+x}{k+x}\right)\right]. \end{aligned}$$

Each summand has two ln-terms; we separate them into two sums and compute each as a telescoping product. Writing out the partial sum to N:

$$\sum_{k=1}^{N} \ln \frac{k+1}{k} = \ln \frac{2}{1} + \ln \frac{3}{2} + \cdots + \ln \frac{N+1}{N} = \ln \frac{2 \cdot 3 \cdots (N+1)}{1 \cdot 2 \cdots N} = \ln(N+1),$$

$$\sum_{k=1}^{N} \ln \frac{k+1+x}{k+x} = \ln \frac{2+x}{1+x} + \ln \frac{3+x}{2+x} + \cdots + \ln \frac{N+1+x}{N+x} = \ln \frac{N+1+x}{1+x}.$$

In both cases, consecutive numerators and denominators cancel, leaving only the first denominator and the last numerator. Subtracting and taking $N \to \infty$:

$$\frac{1}{\ln 2}\left[\ln(N+1) - \ln \tfrac{N+1+x}{1+x}\right] = \frac{1}{\ln 2} \ln \frac{(N+1)(1+x)}{N+1+x} \xrightarrow{N \to \infty} \frac{1}{\ln 2} \ln(1+x),$$

since $(N+1)/(N+1+x) \to 1$. Therefore

$$\mu(T^{-1}([0, x])) = \frac{1}{\ln 2} \ln(1+x) = \mu([0, x]). \qquad \square$$

4. The Gauss–Kuzmin theorem

We have shown that the Gauss measure μ is invariant under T, so a number sampled from μ keeps the same distribution at every step. But a typical number is sampled from Lebesgue measure, not from μ. The question driving this section is: does it matter? If we start with Lebesgue measure and iterate T, does the distribution of the nth remainder $T^n(\alpha)$ eventually converge to the Gauss measure? If so, the long-run frequencies of the partial quotients are determined by μ, and the Gauss–Kuzmin distribution follows. We proceed in three stages: first we derive a functional equation that governs how the distribution evolves from step to step, then we show that it converges exponentially fast to the Gauss measure (the Gauss–Kuzmin–Lévy theorem), and finally we translate this convergence into the frequency formula.

To make the convergence question precise, we track how the distribution of the nth remainder evolves with n. Define the *distribution function* $\mu_n : [0,1] \to [0,1]$ by

$$\mu_n(x) = \mathcal{M}\big(\{\alpha \in (0,1) : T^n(\alpha) < x\}\big).$$

This records the Lebesgue measure of the set of starting points whose nth remainder falls below x. Our goal is to show that $\mu_n(x) \to \log_2(1+x)$ as $n \to \infty$, i.e., that the distribution of the nth remainder converges to the Gauss measure. Once this is established, the frequency formula will follow. Since $x_0 = \alpha$ is uniformly distributed on $(0,1)$ under Lebesgue measure, $\mu_0(x) = x$.

4.1. **The functional equation.** The key tool for analyzing the convergence of μ_n is the following recurrence, which expresses μ_{n+1} in terms of μ_n.

PROPOSITION 4.1. *For all $n \geq 0$ and all $x \in [0,1]$,*

$$\mu_{n+1}(x) = \sum_{k=1}^{\infty} \left[\mu_n\left(\frac{1}{k}\right) - \mu_n\left(\frac{1}{k+x}\right)\right]. \tag{4.1}$$

Proof. The remainder $x_{n+1} = T(x_n)$ is less than x if and only if, for some $k \geq 1$, the previous remainder x_n lies in the kth branch of T and satisfies $1/x_n - k < x$. This is equivalent to $\frac{1}{k+x} < x_n \leq \frac{1}{k}$. Summing over all possible values of k gives (4.1). □

One can verify directly that $\varphi(x) = \log_2(1+x)$ satisfies the fixed-point equation

$$\varphi(x) = \sum_{k=1}^{\infty} \left[\varphi\left(\frac{1}{k}\right) - \varphi\left(\frac{1}{k+x}\right)\right].$$

Indeed, substituting $\varphi(x) = \log_2(1+x)$ reduces the fixed-point equation to the identity

$$\ln(1+x) = \sum_{k=1}^{\infty} \left[\ln\left(1+\frac{1}{k}\right) - \ln\left(1+\frac{1}{k+x}\right)\right],$$

which telescopes exactly as in the proof of Theorem 3.5. Since $\log_2(1+x)$ is a fixed point of the recursion (4.1) and $\mu_0(x) = x$ is not far from it, one expects $\mu_n(x) \to \log_2(1+x)$ as $n \to \infty$.

4.2. **Convergence to the Gauss measure.** The following theorem confirms this convergence and gives an exponential rate. Its history spans a century. Gauss conjectured the convergence $\mu_n(x) \to \log_2(1+x)$ in a letter to Laplace [9], but could not prove it. In 1928, Kuzmin [5] gave the first proof, obtaining the rate $O(e^{-\lambda\sqrt{n}})$. Just a year later, Lévy [6], using methods from probability theory, sharpened the rate to the exponential bound $O(e^{-\lambda n})$ stated below. The proof sketch we give follows Khinchin's presentation [1, §15], which uses Kuzmin-style ideas but obtains Lévy's stronger $O(e^{-\lambda n})$ rate.

THEOREM 4.2 (Gauss–Kuzmin–Lévy [5, 6]). *There exist absolute positive constants $\lambda > 0$ and $A > 0$ such that*

$$|\mu_n(x) - \log_2(1+x)| < A\,e^{-\lambda n}$$

for all $0 \leq x \leq 1$ and all $n \geq 0$.

Proof sketch [1, §15]. The idea is to work with the density $f_n(x) = \mu_n'(x)$ of the nth remainder rather than the distribution function μ_n itself, since the density satisfies a cleaner recursion. Differentiating the functional equation (4.1) gives

$$f_{n+1}(x) = \sum_{k=1}^{\infty} \frac{1}{(k+x)^2}\, f_n\!\left(\frac{1}{k+x}\right). \tag{4.2}$$

The target density is $f_\infty(x) = \frac{1}{\ln 2} \cdot \frac{1}{1+x}$ (the density of the Gauss measure). Since $\mu_0(x) = x$, we start from $f_0(x) = 1$, and we need to show that repeated application of (4.2) drives f_n toward f_∞ exponentially fast.

Step 1: The shape is right from the start. Notice that $f_\infty(x) = \frac{1}{\ln 2}\cdot\frac{1}{1+x}$ has the shape $c/(1+x)$ for a constant c. The key observation is that the recursion (4.2) approximately preserves this shape. If f_n is close to a multiple of $1/(1+x)$, then f_{n+1} is even closer. To make this precise, we track how tightly $f_n(x)$ is sandwiched between two multiples of $1/(1+x)$. Define

$$g_n = \inf_{x\in(0,1)} \big[(1+x)f_n(x)\big], \qquad G_n = \sup_{x\in(0,1)} \big[(1+x)f_n(x)\big].$$

These are the best constants such that $\frac{g_n}{1+x} \le f_n(x) \le \frac{G_n}{1+x}$ for all x. The gap $G_n - g_n$ measures how far f_n is from having the exact shape $c/(1+x)$—when the gap is zero, f_n is a perfect multiple of $1/(1+x)$, and hence equals f_∞ (since f_n integrates to 1).

Step 2: The gap shrinks exponentially. To see where the contraction comes from, substitute the upper sandwich bound $f_n(t) \le G_n/(1+t)$ into the recursion (4.2). The sum $\sum_{k=1}^{\infty} \frac{1}{(k+x)^2} \cdot \frac{G_n}{1+1/(k+x)}$ telescopes (by the same partial-fraction mechanism as in the proof of Theorem 3.5) to $G_n/(1+x)$, so the recursion maps $G_n/(1+x)$ back to itself. More careful bookkeeping, tracking both the upper and lower bounds through the recursion and using the identity $\sum_{k=1}^{\infty} \frac{1}{(k+x)(k+x+1)} = \frac{1}{1+x}$, shows that the new gap satisfies

$$G_{n+1} - g_{n+1} < \delta\,(G_n - g_n) + 2^{-n+2}(\beta + G_n),$$

where $\delta = 1 - \frac{\ln 2}{2} \approx 0.654 < 1$ and the constant β bounds $|f_0'(x)|$. The first term contracts the gap by a factor of $\delta < 1$; the second decays as $2^{-n} \to 0$. Since both terms decay exponentially, $G_n - g_n \to 0$ at an exponential rate, and g_n, G_n converge to the common limit $\frac{1}{\ln 2}$ (forced by $\int_0^1 f_n = 1$). This gives

$$\left| f_n(x) - \frac{1/\ln 2}{1+x} \right| < Be^{-\lambda n}$$

for absolute constants $B, \lambda > 0$.

Step 3: From density to distribution. The bound above is on the density f_n. Integrating from 0 to x recovers the distribution function:

$$\left| \mu_n(x) - \frac{\ln(1+x)}{\ln 2} \right| = \left| \int_0^x \big(f_n(t) - f_\infty(t)\big)\, dt \right| \le \int_0^x Be^{-\lambda n}\, dt \le Be^{-\lambda n},$$

which gives the exponential rate stated in the theorem. (Kuzmin's original argument, which is slightly different, yields only $O(e^{-\lambda\sqrt{n}})$; the cleaner exponential rate is due to Lévy's approach.) □

Remark 4.3. Lévy's bound gives $\lambda = \pi^2/(12\ln 2) \approx 1.186$. Wirsing [7] later determined the optimal constant $\lambda_W \approx 0.3037$, now called the *Gauss–Kuzmin–Wirsing constant*, which governs the true rate of convergence to the Gauss measure. Any of these rates is more than sufficient for the corollaries that follow.

4.3. **The frequency formula.** The Gauss–Kuzmin–Lévy theorem determines the distribution of the iterates $T^n(\alpha)$. Since the event $a_{n+1}(\alpha) = k$ corresponds to $T^n(\alpha)$ lying in the interval $\big(1/(k+1),\, 1/k\big]$, the single-time distribution of each partial quotient follows directly:

$$\mathcal{M}\big(\{\alpha : a_{n+1}(\alpha) = k\}\big) = \mu_n\left(\frac{1}{k}\right) - \mu_n\left(\frac{1}{k+1}\right) \xrightarrow{n\to\infty} \log_2\left(1 + \frac{1}{k(k+2)}\right) = d(k).$$

From distribution to frequency: why ergodicity is needed. We should pause and ask: what exactly has been proved? The calculation above shows that if we pick α at random, then for large n the partial quotient a_{n+1} equals k with probability $\approx d(k)$. This is an average *across* many α's at one fixed time step. We want something stronger: fix a single α and count the proportion of its own partial quotients equal to k along the orbit $\alpha, T(\alpha), T^2(\alpha), \ldots$ These two kinds of average—across space vs. along an orbit—need not agree in general. The question is whether the Gauss map "mixes" $(0,1)$ thoroughly enough that the time average along a single orbit matches the space average.

This is precisely the role of *ergodicity*. The Gauss map T is ergodic with respect to μ: every T-invariant measurable set has μ-measure 0 or 1, so orbits cannot "get stuck" in a proper subset of $(0,1)$. Once ergodicity is established, the Birkhoff ergodic theorem [8] guarantees that time averages along almost every orbit equal the corresponding space averages.

Two ingredients are used as black boxes: the ergodicity of T with respect to μ, and the following classical theorem of Birkhoff, which converts ergodicity into a concrete statement about time averages.

Theorem 4.4 (Birkhoff's ergodic theorem [8]). *Let T be a measure-preserving transformation of a probability space (X,μ), and suppose T is ergodic. Then for any μ-integrable function $f : X \to \mathbb{R}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) = \int_X f\, d\mu$$

for μ-almost every $x \in X$. In words, the time average of f along almost every orbit equals the space average. For a proof, see [3, Chapter 2].

Corollary 4.5 (Gauss–Kuzmin distribution [1, Chapter III, §15]). *For almost every $\alpha \in (0,1)$,*

$$\lim_{N\to\infty} \frac{1}{N} \#\{1 \le n \le N : a_n(\alpha) = k\} = \log_2\left(1 + \frac{1}{k(k+2)}\right)$$

for every positive integer k.

Proof. For a fixed positive integer k, let $f = \mathbf{1}_{(1/(k+1),\,1/k]}$ be the indicator function of the interval where $a_1 = k$. Since T is ergodic and measure-preserving with respect to μ, Theorem 4.4 gives, for μ-almost every α,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(\alpha)) = \int_0^1 f\, d\mu = \mu\big((1/(k+1), 1/k]\big).$$

The left-hand side counts the proportion of iterates $T^n(\alpha)$ that land in $(1/(k+1), 1/k]$, which is exactly the proportion of partial quotients $a_{n+1}(\alpha)$ equal to k. Computing the right-hand side:

$$\mu\big((1/(k+1), 1/k]\big) = \frac{1}{\ln 2}\int_{1/(k+1)}^{1/k} \frac{dx}{1+x} = \frac{1}{\ln 2}\left[\ln\!\big(1+\tfrac{1}{k}\big) - \ln\!\Big(1+\tfrac{1}{k+1}\Big)\right]$$
$$= \log_2\left(1 + \frac{1}{k(k+2)}\right).$$

Since the Gauss measure μ is absolutely continuous with respect to Lebesgue measure with density $\frac{1}{(\ln 2)(1+x)}$ (which is bounded above and below on $(0,1)$), a μ-null set is also Lebesgue-null, so the conclusion holds for Lebesgue-almost every α as well. ☐

The ergodicity of T with respect to μ is itself a nontrivial fact. It can be proved using the Knopp lemma or by applying the Gauss–Kuzmin–Lévy theorem to indicator functions of fundamental intervals; see [3, Section 3.2] for an exposition.

5. Khinchin's theorem and Khinchin's constant

The Gauss–Kuzmin distribution tells us the frequency of each individual partial quotient value. A natural next question is what we can say about averages and products of partial quotients? Since we know the limiting frequency of each value m, we should be able to compute the long-run average of any reasonable function $f(a_j)$ by weighting $f(m)$ by its frequency $d(m)$. This is the content of Khinchin's theorem, whose most famous application is the existence of a universal constant governing the geometric mean of the partial quotients.

Theorem 5.1 (Khinchin [1, Chapter III, §16]). *Let $f : \mathbb{N} \to (0, \infty)$ be a function of a positive integer. Suppose there exist positive constants C and δ such that $f(m) < Cm^{1/2-\delta}$ for $m = 1, 2, \ldots$ Then, for almost every $\alpha \in (0, 1)$,*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} f(a_j) = \sum_{m=1}^{\infty} f(m)\, \log_2\left(1 + \frac{1}{m(m+2)}\right). \tag{5.1}$$

Proof sketch. The idea is to reduce Khinchin's theorem to Birkhoff's ergodic theorem (Theorem 4.4). Define $g : (0,1) \to \mathbb{R}$ by $g(x) = f(\lfloor 1/x \rfloor)$, so that $g(\alpha) = f(a_1(\alpha))$ and more generally $g(T^{j-1}(\alpha)) = f(a_j(\alpha))$. Then the left-hand side of (5.1) is exactly $\frac{1}{N}\sum_{j=0}^{N-1} g(T^j(\alpha))$, which is the time average of g along the orbit of α.

By Theorem 4.4, this time average converges to the space average $\int_0^1 g\, d\mu$ for μ-almost every α, provided g is μ-integrable. Now

$$\int_0^1 g\, d\mu = \sum_{m=1}^{\infty} \int_{1/(m+1)}^{1/m} f(m)\, \frac{1}{\ln 2} \cdot \frac{dx}{1+x} = \sum_{m=1}^{\infty} f(m)\, \log_2\left(1 + \frac{1}{m(m+2)}\right),$$

since g is constant equal to $f(m)$ on $(1/(m+1), 1/m]$. The growth condition $f(m) < Cm^{1/2-\delta}$ ensures this series converges (since $d(m) \sim 1/(m^2 \ln 2)$, we get $f(m)\, d(m) = O(m^{-3/2-\delta})$), guaranteeing the μ-integrability of g. The conclusion transfers from μ-almost-everywhere to Lebesgue-almost-everywhere by mutual absolute continuity. ☐

5.1. **Khinchin's constant.** The most famous application of Theorem 5.1 is the case $f(m) = \ln m$. Since $\ln m < m^{1/2-\delta}$ for any $\delta > 0$ and large m, the growth condition is satisfied. By Theorem 5.1, for almost every α,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \ln a_j = \sum_{m=1}^{\infty} \ln(m)\, \log_2\left(1 + \frac{1}{m(m+2)}\right). \tag{5.2}$$

The left-hand side is $\frac{1}{N}\ln(a_1 a_2 \cdots a_N) = \ln(a_1 a_2 \cdots a_N)^{1/N}$, so exponentiating both sides and writing $\exp(\sum \ln(m)\cdot c_m) = \prod m^{c_m}$ gives:

Corollary 5.2 (Khinchin's constant). *For almost every $\alpha \in (0,1)$,*

$$\lim_{N\to\infty} (a_1 a_2 \cdots a_N)^{1/N} = \prod_{m=1}^{\infty} m^{\log_2(1+1/(m(m+2)))} =: K_0. \tag{5.3}$$

The constant K_0 is called Khinchin's constant, *and its numerical value is $K_0 =$ $2.685\,452\,001\,065\,306\,445\ldots$*

Corollary 5.2 is remarkable: for almost every real number, the geometric mean of the first N partial quotients converges to the same universal constant K_0, regardless of the number.

Remark 5.3 (Arithmetic mean). The arithmetic mean $(a_1+\cdots+a_N)/N$ does not converge for almost every α. The expected value $\sum_{m=1}^{\infty} m \cdot d(m)$ diverges because the function $f(m) = m$ does not satisfy the growth condition of Theorem 5.1. In fact, for almost every α, the partial quotients are unbounded (see [1, §13]), and these occasional large values are enough to force the arithmetic mean to diverge.

References

[1] A. Ya. Khinchin, *Continued Fractions*, University of Chicago Press, 1964.

[2] J. Borwein, A. van der Poorten, J. Shallit, and W. Zudilin, *Neverending Fractions: An Introduction to Continued Fractions*, Cambridge University Press, 2014.

[3] M. Einsiedler and T. Ward, *Ergodic Theory with a View towards Number Theory*, Graduate Texts in Mathematics, vol. 259, Springer, London, 2011.

[4] T. Tao, *An Introduction to Measure Theory*, Graduate Studies in Mathematics, vol. 126, American Mathematical Society, Providence, RI, 2011.

[5] R. O. Kuzmin, "On a problem of Gauss," *Dokl. Akad. Nauk SSSR* Ser. A (1928), 375–380.

[6] P. Lévy, "Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue," *Bull. Soc. Math. France* **57** (1929), 178–194.

[7] E. Wirsing, "On the theorem of Gauss–Kuzmin–Lévy and a Frobenius-type theorem for function spaces," *Acta Arith.* **24** (1974), 507–528.

[8] G. D. Birkhoff, "Proof of the ergodic theorem," *Proc. Natl. Acad. Sci. USA* **17** (1931), 656–660.

[9] C. F. Gauss, "Problema e calculo probabilitatis circa fractiones continuas," in *Werke*, Bd. X/1, Königliche Gesellschaft der Wissenschaften, Göttingen, 1917, pp. 552–556. See also the letter to Laplace of 30 January 1812, ibid., p. 371.

[10] P. Sibbertsen, T. Lampert, K. Müller, and M. Taktikos, "Do algebraic numbers follow Khinchin's Law?" Hannover Economic Papers No. 686, Leibniz Universität Hannover, 2021.

[11] P. G. L. Dirichlet, "Verallgemeinerung eines Satzes aus der Lehre von den Kettenbrüchen nebst einigen Anwendungen auf die Theorie der Zahlen," *S. B. Preuss. Akad. Wiss.* (1842), 93–95.

[12] A. Hurwitz, "Ueber die angenäherte Darstellung der Irrationalzahlen durch rationale Brüche," *Math. Ann.* **39** (1891), 279–284.

Massachusetts Institute of Technology
Email address: `fionalee@mit.edu`