Kuzmin’s Zero Measure Extraordinary Set

Matthew Michelini
Mathematics Department
Princeton University

April 9, 2004

My special thanks to my parents who have made my Princeton experience possible. This thesis and more importantly graduation would not have been possible if not for the guidance, patience, and unfaltering faith of both Ramin Takloo-Bighash and Steven Miller. Because of Dr. Takloo-Bighash and Dr. Miller, I conquered the impossible.

Contents

Abstract

In this paper, I will reconstruct Khintchine's presentation of Kuzmin's Theorem but with vastly more details and explanations. I will then use this formulation to give a method of approximating the absolute positive constants A and λ in Levy's error term:

$$\left| \mu(E\binom{n}{k}) - \frac{\ln\left\{1 + \frac{1}{k(k+2)}\right\}}{\ln 2} \right| < \frac{A}{k(k+1)} e^{-\lambda(n-1)}.$$

I conducted a numerical experiment to estimate these constants given certain conditions and will present the results in this paper.

Finally, there exists some guiding theory to describe the zero measure set of $\alpha \in [0,1]$ that does not obey Kuzmin's Theorem, but the existing theory has never been fully summarized in a single exposition. I will provide this summary, and I will present two additional sets, for which theory suggests Kuzmin's Theorem does not hold.

Chapter 1

Preliminaries

1.1 Notation

A continued fraction is the representation of a number $\alpha \in \mathbb{R}$ and is of the form:

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}} \tag{1.1.1}$$

If the continued fraction is infinite, then the expansion will not terminate with a_n like the expansion above does. Throughout this thesis we will also represent a finite continued fraction with $[a_0; a_1, \ldots, a_n]$ and an infinite continued fraction with $[a_0; a_1, \ldots]$.

Where appropriate, I will make the following notational distinctions: $x = [a_0; a_1, \ldots]$ is the value of the continued fraction of arbitrary length, but α is the actual number being represented by the continued fraction, or the number to which the continued fraction x converges. In other words, we attempt to represent α by a rational number, $x = [a_0; a_1, \ldots]$. For example, if $\alpha = \pi$, then $x = [3; 7, 15, 1, 292, 1, \ldots]$ is the value of the continued fraction expansion of arbitrary length that ultimately converges to π.

1.2 Definitions

Many of the following definitions are found in [MT]:

Definition 1.2.1. ***Coefficients of a Continued Fraction:*** *If $x = [a_0; a_1, \ldots]$, then the a_i are the digits or coefficients.*

Definition 1.2.2. ***Positive Continued Fraction:*** *A continued fraction $[a_0; a_1, \ldots]$ is positive if each $a_i > 0$.*

Definition 1.2.3. ***Simple Continued Fraction:*** *A continued fraction is simple if all a_i are positive integers.*

Definition 1.2.4. ***Convergents of a Continued Fraction:*** *Let $x = [a_0; a_1, \ldots]$, and if $x_n = [a_0; a_1, \ldots, a_n] = \frac{p_n}{q_n}$, then $\frac{p_n}{q_n}$ is the n^{th} quotient or convergent.*

Property 1.2.5. *Let $k \geq 2$, then the following is an increasing sequence for even values of k and a decreasing for odd values of k:*

$$\frac{p_{k-2}}{q_{k-2}}, \frac{p_{k-2} + p_{k-1}}{q_{k-2} + q_{k-1}}, \frac{p_{k-2} + 2p_{k-1}}{q_{k-2} + 2q_{k-1}}, \ldots, \frac{p_{k-2} + a_k p_{k-1}}{q_{k-2} + a_k q_{k-1}} = \frac{p_k}{q_k} \tag{1.2.2}$$

Definition 1.2.6. ***Intermediate Fractions:*** *The fractions standing between $\frac{p_{k-2}}{q_{k-2}}$ and $\frac{p_k}{q_k}$ are called intermediate fractions.*

Let us define:

$$s_k = [a_0; a_1, \ldots, a_k], \tag{1.2.3}$$

or a section comprised of the first k coefficients of a continued fraction. Correspondingly, r_k is the remainder of the continued fraction beginning with the coefficient a_k:

$$r_k = [a_k; a_{k+1}, \ldots], \tag{1.2.4}$$

where r_k terminates at a_n if the continued fraction expansion is finite, or $x = [a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_{k-1}, r_k]$. If r_k exists, then $r_k \geq 1$ because $a_k \geq 1$ for all k.

1.3 Finding the Quotients of a Continued Fraction and Uniqueness

Although there exist other methods to determine a continued fraction's coefficient values, these algorithms build on the same idea behind the Lang-Trotter method [LT]. However, the "old-fashioned" method utilizes the Euclidean algorithm and implementing this method yields x's unique continued fraction expansion [MT].

Assume that x is represented by a simple continued fraction:

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}} \tag{1.3.5}$$

To find a_0, we take the greatest integer less than x (denoted by $[x]$) to be a_0's value. Hence, the remainder of the continued fraction expansion represents the value $x - [x]$:

$$x - [x] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}, \tag{1.3.6}$$

where by taking the reciprocal, we determine a_1's value:

$$\begin{aligned} x_1 = \frac{1}{x - [x]} &= a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}} \\ \Rightarrow [x_1] &= a_1. \end{aligned} \tag{1.3.7}$$

We repeat this process for x_2 $\left(\text{i.e. } x_2 = \frac{1}{x_1 - [x_1]}\right)$ and all subsequent x_i until our expansion repeats, terminates, or a desired n is reached.

Since we assumed that x's continued fraction representation was simple, the only modification needed if $x < 0$ is to allow $a_0 < 0$, then every other coefficient value is a natural number. Assuming that a continued fraction possesses at least m coefficients, it is also important to reiterate that $a_i \geq 1$ for $1 \leq i \leq m$ because $0 \leq x_i - [x_i] < 1$. If $x_i - [x_i] = 0$ then the remainder is 0 and the expansion terminates.

From this traditional method of finding coefficients, we note a very important property:

Property 1.3.1. *Let α be a rational number with a finite continued fraction expansion of length n, then $\alpha = [a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_{k-1}] + \frac{1}{r_k}$, where a_i is a positive integer for all $i \leq n$.*

For expositions on this see [MT]. We will now show that if x has an infinite continued fraction expansion, then the expansion is unique.

Theorem 1.3.2. *Let $x = [a_0; a_1, a_2, a_3, \ldots] = [a_0^{'}; a_1^{'}, a_2^{'}, a_3^{'}, \ldots]$ be two continued fraction expansions for x, then $a_i = a_i^{'}$ for all i. Thus, a continued fraction expansion is unique.*

Proof: This proof is found in [Ki]. Let the conditions of the theorem hold, namely $x = [a_0; a_1, a_2, a_3, \ldots] = [a_0^{'}; a_1^{'}, a_2^{'}, a_3^{'}, \ldots]$, where the expansions can be finite or infinite. Let $i = 0$, then since both expansions represent x, we have $a_0 = [x]$ and $a'_0 = [x]$, which implies $a_0 = a_0^{'}$. Assume $a_i = a_i^{'}$ for all $i \leq n$. Then by Theorem 1.4.2, we have $p_i = p_i^{'}$ and $q_i = q_i^{'}$ for all $i \leq n$. From the definition of r_k and Theorem 1.4.8, we have $x = [a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_{k-1}, r_k]$, which implies:

$$x = \frac{p_n r_{n+1} + p_{n-1}}{q_n r_{n+1} + q_{n-1}} = \frac{p_n^{'} r_{n+1}^{'} + p_{n-1}^{'}}{q_n^{'} r_{n+1}^{'} + q_{n-1}^{'}} = \frac{p_n r_{n+1}^{'} + p_{n-1}}{q_n r_{n+1}^{'} + q_{n-1}}; \tag{1.3.8}$$

therefore, $r_{n+1} = r_{n+1}^{'}$. But we know that $a_{n+1} = [r_{n+1}]$ and $a_{n+1}^{'} = [r_{n+1}^{'}]$ so $a_{n+1} = a_{n+1}^{'}$, and the two expansions are identical.□

From this proof we conclude if α's continued fraction expansion terminates with a coefficient $a_n = 1$, then two possible continued fraction representations exist for α: one representation ends with $a_n = 1$ and the other ends with $a_{n-1}^{'} = a_{n-1} + 1$. Only in this case is it possible for a number to have two distinct continued fraction representations.

1.4 Properties of Convergents

Property 1.4.1. *Let $[a_0; a_1, \ldots, a_n]$ be a continued fraction, then [MT]:*

$$\begin{aligned} 1.\ [a_0; a_1, \ldots, a_n] &= [a_0; a_1, \ldots, a_{n-1} + \frac{1}{a_n}] \\ 2.\ [a_0; a_1, \ldots, a_n] &= [a_0; a_1, \ldots, a_{m-1}, [a_m, \ldots, a_n]]. \end{aligned} \tag{1.4.9}$$

The following theorem will be integral in answering certain questions involving measure and continued fractions, especially with regard to a continued fraction's tail or interval of uncertainty.

Theorem 1.4.2. *For any* $m \in \{2, \ldots, n\}$ *and* a_m *a positive integer, we have*

$$\begin{aligned} 1.\ p_0 = a_0,\ p_1 = a_0a_1 + 1, \qquad &\textit{and}\ \ p_m = a_m p_{m-1} + p_{m-2} \\ 2.\ q_0 = 1,\ \ q_1 = a_1, \qquad &\textit{and}\ \ q_m = a_m q_{m-1} + q_{m-2}. \end{aligned} \tag{1.4.10}$$

Proof: Use induction (see [MT]). Note p_n and q_n are positive integers, and we shall assume this when referencing them hereinafter.

This theorem provides a closed form expression for the denominator, the numerator, and the convergents of the continued fraction expansion of $\alpha \in \mathbb{R}$. In fact, we will soon understand that this formula allows us to estimate how well x approximates α given N coefficients.

Lemma 1.4.3. *For all* $k \geq 0$ *we have* $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$.

Lemma 1.4.4. *For all* $k \geq 1$ *we have* $p_n q_{n-2} - p_{n-2} q_n = (-1)^n a_n$.

The next theorem is presented only to justify later the assumption $(p, q) = 1$:

Theorem 1.4.5. *If a continued fraction expansion possesses an* n^{th} *convergent* $\frac{p_n}{q_n}$*, then* $\frac{p_n}{q_n}$ *is reduced.*

Proof: From Lemma 1.4.3 any common factor c of both p_n and q_n is also a factor of $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$. So $c|(-1)^{n-1}$ which implies $c = 1$. □

This theorem implies the reduced value of $x = [a_0; a_1, \ldots, a_k]$ is given by the k^{th} convergent $\frac{p_k}{q_k}$. In the case the continued fraction expansion is finite, there exist some n, for which $\alpha = \frac{p_n}{q_n} = x$.

Lemma 1.4.6. *For all* $k \geq 2$ *we have*

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n-1}}{q_n q_{n-1}} \tag{1.4.11}$$

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{(-1)^n a_n}{q_n q_{n-2}}. \tag{1.4.12}$$

Proof: We know from Theorem 1.4.2 that $q_i > 0$ for all i, otherwise the convergent would be undefined. Moreover, $a_i > 0$ for all i, otherwise the expansion would terminate. Therefore, we can divide Lemma 1.4.3 by $q_n q_{n-1}$, by which the first relation of Lemma 1.4.6 follows. To obtain the second relation, we divide Lemma 1.4.4 by $q_n q_{n-2}$. □

Given a continued fraction representation of length n, there exsits an **interval of uncertainty**, which is the absolute difference between $\frac{p_n}{q_n}$ and α. Lemma 1.4.6 offers a closed form expression for the measure (or length) of the interval of uncertainty given a continued fraction expansion of length n. Additionally, this lemma, combined with Theorem 1.4.2, allows us to determine how fast the measure of the interval of uncertainty $|\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}}|$ falls as n becomes large. The next theorem will show the measure of the interval of uncertainty decreases after each coefficient.

Theorem 1.4.7. *If the coefficients a_0 to a_n are positive, then the sequence x_{2m} is an increasing sequence, the sequence x_{2m+1} is a decreasing sequence, and for every m, $x_{2m} < \alpha < x_{2m+1}$ (if $n \neq 2m$ or $2m+1$) and where $x_m = \frac{p_m}{q_m}$.*

Proof: This proof is adapted from [MT]. We will prove this theorem for α irrational. By Lemma 1.4.6 we know that

$$x_{2m+2} - x_{2m} = \frac{(-1)^{2m} a_{2m}}{q_{2m} q_{2m+2}} > 0. \tag{1.4.13}$$

This equation holds for all m because $a_m > 0$ and the continued fraction expansion of an irrational number is infinite. Since the right hand side is always positive, this implies that $x_{2m+2} > x_{2m}$. A similar relationship can be proved for the odd indexed terms but the right hand side of Equation 1.4.13 will have the factor $(-1)^{2m+1}$ instead of $(-1)^{2m}$. Since -1 is raised to an odd power for the odd indexed terms, the difference $x_{2m+3} - x_{2m+1} < 0 \quad \Rightarrow \quad x_{2m+1} > x_{2m+3}$. We must show now that $x_{2m+1} - x_{2m} > 0$, which is just another application of Lemma 1.4.6 taking $n^{'} = 2m+1$. If $n^{'}$ is odd, then $x_{n^{'}} = \frac{p_{n^{'}}}{q_{n^{'}}}$ is less than all the preceeding odd indexed convergents. If $n^{'}$ is even, then $x_{n^{'}} = \frac{p_{n^{'}}}{q_{n^{'}}}$ is greater than all the preceeding even indexed convergents. Since $x_n \geq x_{2m}$ for even n, $x_n \leq x_{2m+1}$ for odd n, and $x_{2m+1} > x_{2m}$ for all m, we conclude $x_n = \frac{p_n}{q_n}$ ($n \geq 2m+1$) will always be in the interval $[x_{2m}, x_{2m+1}]$ for all n. □

The next few theorems will be used extensively in later chapters.

Theorem 1.4.8. *For any* k $(1 \leq k \leq n)$ *we have:*

$$\alpha = [a_0; a_1, \ldots, a_n] = \frac{p_{k-1}r_k + p_{k-2}}{q_{k-1}r_k + q_{k-2}}. \tag{1.4.14}$$

Proof: Recall our definition of r_k implied $[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_{k-1}, r_k]$, where we assume that $\frac{p_{k-1}}{q_{k-1}}$ is the $(k-1)^{th}$ convergent of the continued fraction on the right hand side of Equation 1.4.14, but this continued fraction's k^{th} order convergent's value is α and given by Theorem 1.4.2:

$$p_k = p_{k-1}r_k + p_{k-2} \;\; and \;\; q_k = q_{k-1}r_k + q_{k-2}. \;\; \square \tag{1.4.15}$$

Theorem 1.4.9. *Let* $[a_0; a_1, \ldots, a_n]$ *be a positive, simple continued fraction. Then:*

$$\begin{aligned} &1.\ q_n \geq q_{n-1} \;\; \textit{for all} \;\; n \geq 1, \;\; \textit{and} \;\; q_n > q_{n-1} \;\; \textit{if} \;\; n > 1. \\ &2.\ q_n \geq n, \;\; \textit{with strict inequality if} \;\; n > 3. \end{aligned} \tag{1.4.16}$$

Note, a similar statement can be made for the numerators (p_n's) of the convergents.

Proof: From Theorem 1.4.2 we know $q_0 = 1$, $q_1 = a_1 \geq 1$, and $q_n = a_n q_{n-1} + q_{n-2}$. For all n, we have $a_n \geq 1$ and $a_n \in \mathbb{N}$. Thus, $a_n q_{n-1} + q_{n-2} = q_n \geq q_{n-1}$. If $n > 1$ and $q_{n-2} > 0$, then the inequality is strict.

The second claim is proved by induction. For $n = 0$, the claim is clearly satisfied as $q_0 = 1 > 0$. Assume $q_{n-1} \geq n - 1$ for all $i < n$. Then from Equation 1.4.2, we have $q_n = a_n q_{n-1} + q_{n-2} \geq q_{n-1} + q_{n-2} \geq (n-1) + 1 = n$, where the last inequality is given by the inductive step. If at any point the inequality is strict, then it is strict from that point onward. By inspection, it is easy to see that $q_n > n$ for $n > 3$ because $q_n = a_n q_{n-1} + q_{n-2}$; letting $n = 4$ and $a_n = 1$, we have $q_4 = 3 + 2 = 5 > 4 = n$. $\square$

The ratio $\frac{q_k}{q_{k-1}}$ is often needed in establishing convergence properties of continued fractions.

Theorem 1.4.10. *For any* $k \geq 1$, $x = [a_0; a_1, \ldots, a_k, \ldots]$, *and when convergents* $\frac{p_{k-1}}{q_{k-1}}, \frac{p_k}{q_k}$ *exist, we have:*

$$\frac{q_k}{q_{k-1}} = [a_k; a_{k-1}, \ldots, a_1]. \tag{1.4.17}$$

Proof: This theorem is easily proved by induction. Let $k = 1$, then the theorem is trivially true because $\frac{q_1}{q_0} = \frac{a_1}{1} = a_1$. Now assume $\frac{q_{k-1}}{q_{k-2}} = [a_{k-1}; a_{k-2}, \ldots, a_1]$ holds for all $i < k$. By combining the relation $q_k = a_k q_{k-1} + q_{k-2}$ established in Theorem 1.4.2 with Property 1.3.1, we reason:

$$\frac{q_k}{q_{k-1}} = a_k + \frac{q_{k-2}}{q_{k-1}} = [a_k; \frac{q_{k-1}}{q_{k-2}}], \tag{1.4.18}$$

where $\frac{q_{k-2}}{q_{k-1}}$ is equivalent to term $\frac{1}{r}$ in Property 1.3.1. By Definition 1.2.4 and our inductive assumption, we conclude:

$$\begin{aligned} [a_k; \frac{q_{k-1}}{q_{k-2}}] &= a_k + [0; a_{k-1}, \ldots, a_1] \\ \Rightarrow \frac{q_k}{q_{k-1}} &= [a_k; a_{k-1}, \ldots, a_1]. \quad \square \end{aligned} \tag{1.4.19}$$

We note relationship in Theorem 1.4.10 is simply the reverse of $x = [a_0; a_1, \ldots, a_k]$ excluding the coefficient a_0.

1.5 Infinite Continued Fractions

We previously defined an infinite continued fraction as $x = [a_0; a_1, \ldots]$, where the expansion never terminates. An infinite continued fraction converges to a value α only if the following limit exists:

$$\lim_{n\to\infty} \frac{p_n}{q_n} = \alpha < \infty \tag{1.5.20}$$

In other words, associated with each coefficient a_n in the above continued fraction is a convergent $\frac{p_n}{q_n}$, and if the sequence of convergents $\frac{p_0}{q_0}, \ldots, \frac{p_n}{q_n}, \ldots$ converges, then the infinite continued fraction converges to value α. However, if the sequence diverges, then the expansion does not converge to a value.

An analog of Theorem 1.4.8 exists for infinite continued fractions, in that we can represent α as follows:

$$\alpha = \frac{p_{n-1} r_n + p_{n-2}}{q_{n-1} r_n + q_{n-2}} \tag{1.5.21}$$

assuming both r_n converges as $k \to \infty$ and q_i, p_i are positive integers for all $i \geq 0$. Recalling our definition of r_k and Theorem 1.4.8 we can write for irrational α:

$$\alpha = [a_0; a_1, \ldots, r_k] = \frac{p_{k-1} \frac{p_k'}{q_k'} - p_{k-2}}{q_{k-1} \frac{p_k'}{q_k'} - q_{k-2}}, \tag{1.5.22}$$

where the right hand side of the equation is the k^{th} order convergent and $r_k = \frac{p_k^{'}}{q_k^{'}}$. If $\lim_{k\to\infty} \frac{p_k^{'}}{q_k^{'}}$ exists, then r_k converges.

We proved in Theorem 1.4.7 that the value of a convergent of an infinite continued fraction is greater than that of any even order convergent and less than that of any odd order convergent. This result will help us to prove later that the convergents of an infinite continued fraction satisfy $\left|\alpha - \frac{p_k}{q_k}\right| < \frac{1}{q_k q_{k-1}}$. For the time being, let us assume this inequality (the proof will come shortly) in order to present the following theorem, which is powerful because it justifies using continued fractions to represent uniquely any $\alpha \in \mathbb{R}$.

Theorem 1.5.1. *To every real number α there corresponds, uniquely, a simple continued fraction whose value is this number. This continued fraction terminates if α is rational, or is infinite if α is irrational [Ki].*

Proof: We established $\alpha = [a_0; r_1] = [a_0; a_1, \ldots, a_n, r_{n+1}]$ in Equation 1.2.4, where r_i is not assumed to be an integer. Let $a_0 = [\alpha]$, where $[x]$ denotes the largest integer not exceeding x. Then by the discussion on finding coefficients, we have $\alpha = a_0 + \frac{1}{r_1}$, where $r_1 = [a_1; a_2, \ldots] > 1$ (equality is not possible unless $[\alpha] = a_0 + 1$). So we have $\frac{1}{r_1} = \alpha - a_0 < 1$; therefore, a_n is the largest integer not exceeding r_n:

$$r_n = a_n + \frac{1}{r_{n+1}} \tag{1.5.23}$$

This process can be repeated indefinitely or until the expansion terminates.

If $\alpha \in \mathbb{Q}$, then $r_n \in \mathbb{Q}$ for all n, and our process will eventually terminate after a finite number of steps. To see this, consider the following: assume that $\alpha \in \mathbb{Q}$, which implies that $r_n = \frac{a}{b}$, then we have

$$r_n - a_n = \frac{a - ba_n}{b} = \frac{c}{b} \tag{1.5.24}$$

where $c < b$ is a strict inequality because $r_n - a_n < 1$. Equation 1.5.23 yields $r_{n+1} = \frac{b}{c}$, but if $c = 0$ then r_n is an integer, in which case the expansion would terminate; so assume that $c \neq 0$. Then $c < b$ and r_{n+1} has a smaller denominator than r_n; as a result, after a finite number of steps in the sequence $r_1, r_2, \ldots$, we must arrive eventually at $r_n = a_n$. Hence, the continued fraction expansion terminates with $a_n = r_n > 1$, where the inequality is strict because $c < b$. By the fact that $\alpha = [a_0; r_1] = [a_0; a_1, \ldots, a_n, r_{n+1}]$, we can conclude that α is indeed represented by a terminating continued fraction. Uniqueness of the terminating continued fraction was proved in Theorem 1.3.2.

If α is irrational, then r_n is irrational for all n, and the process delineated in Equation 1.5.24 never terminates. We write $[a_0; a_1, \ldots, a_n] = \frac{p_n}{q_n}$, where by Theorem 1.4.5, $(p_n, q_n) = 1$. From Theorem 1.4.8 for irrational α, we have:

$$\begin{aligned}
\alpha &= \frac{p_{n-1}r_n + p_{n-2}}{q_{n-1}r_n + q_{n-2}} \\
\text{but} \quad \frac{p_n}{q_n} &= \frac{p_{n-1}a_n + p_{n-2}}{q_{n-1}a_n + q_{n-2}} \\
\text{whence} \quad \alpha - \frac{p_n}{q_n} &= \frac{(p_{n-1}q_{n-2} - q_{n-1}p_{n-2})(r_n - a_n)}{(q_{n-1}r_n + q_{n-2})(q_{n-1}a_n + q_{n-2})} \\
\Rightarrow \left|\alpha - \frac{p_n}{q_n}\right| &< \frac{1}{(q_{n-1}r_n + q_{n-2})(q_{n-1}a_n + q_{n-2})} < \frac{1}{q_n^2}. \quad (1.5.25)
\end{aligned}$$

Because $\frac{1}{q_n^2} \to 0$ as $n \to \infty$, we have $\frac{p_n}{q_n} \to \alpha$, which implies the value of the infinite continued fraction is α. Again, we previously showed uniqueness in Theorem 1.3.2. □

This theorem is important because it enables us to distinguish rational numbers from irrational numbers simply by examing the lengths of the their continued fraction expansions. We concluded in Theorem 1.3.2 that the continued fraction expansion of an irrational number is unique. Therefore, when empirically analyzing properties of a continued fraction expansion, we need not worry other possible results exist.

In summary, every number α can be represented uniquely as a continued fraction. If $\alpha \in \mathbb{Q}$, then the continued fraction expansion terminates; on the other hand, if α is irrational, then the continued fraction expansion is infinite.

The next theorem is a natural consequence of the definition of an infinite continued fraction's value.

Theorem 1.5.2. *If an infinite continued fraction converges, then all its remainders converge. Conversely, if at least one remainder of the infinited continued fraction converges, then the continued fraction itself converges.*

Proof: This proof is an application of Equation 1.5.20 coupled with Theorems 1.4.2 and 1.4.8. For details see [Ki].

A final result will be sufficient to build a strong intuition for infinite continued fractions.

Theorem 1.5.3. *For a simple infinite continued fraction* $[a_0; a_1, \ldots]$ *to converge, it is necessary and sufficient that the series* $\sum_{n=1}^{\infty} a_n$ *should diverge.*

I will present the proof of this theorem but will not consider the question of convergence of an infinite continued fraction when the series $\sum_{n=1}^{\infty} a_n$ converges; however, the argument is worth reading in Khintchine's exposition [Ki].

Proof: As a result of Theorem 1.4.7 and Equation 1.5.20, we note:

$$\lim_{m\to\infty} \frac{p_{2m}}{q_{2m}} = \lim_{m\to\infty} \frac{p_{2m+1}}{q_{2m+1}} = \alpha \tag{1.5.26}$$

is a necessary and sufficient condition to guarantee convergence of an infinite continued fraction to α. Recall from Lemma 1.4.6 that for all $k \geq 1$, we had $\frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k} = \frac{(-1)^k}{q_k q_{k-1}}$; thus, an infinite continued fraction converges if:

$$q_k q_{k-1} \to \infty \text{ as } k \to \infty. \tag{1.5.27}$$

This condition is both necessary and sufficient to guarantee an infinite continued fraction's convergence. Now, we will show if the series $\sum_{n=1}^{\infty} a_n$ diverges, Equation 1.5.27 is satisfied. We label $c = \min[q_0, q_1]$, which implies $q_k \geq c$ for all nonnegative k because $q_k > q_{k-2}$ for all $k \geq 2$ by Theorem 1.4.9, and recall $q_k = q_{k-1}a_k + q_{k-2}$ by Theorem 1.4.2. Collecting our results yields:

$$\begin{aligned} q_k \geq q_{k-2} + ca_k \quad (k \geq 2) \\ \Rightarrow q_{2k} \geq q_0 + c\sum_{n=1}^{k} a_{2n} \\ \text{and } \; q_{2k+1} \geq q_1 + c\sum_{n=1}^{k} a_{2n+1} \\ \Rightarrow q_{2k} + q_{2k+1} \geq q_0 + q_1 + c\sum_{n=1}^{2k+1} a_n \\ \Rightarrow q_k + q_{k-1} > c\sum_{n=1}^{k} a_n, \end{aligned} \tag{1.5.28}$$

where the last inequality holds for all $k \geq 0$. Line 2 is a repeated application of line 1, and line 4 is the sum of lines 2 and 3. Finally, line 5 results from recognizing $q_0, q_1 > 0$. From this last equation, we note at least one of the factors in the product $q_k q_{k-1}$ is greater than $\frac{c}{2}\sum_{n=1}^{k} a_n$, but by line 1 of Equation 1.5.28 the other factor cannot be less than c; so we have for all k,

$$q_k q_{k-1} > \frac{c^2}{2}\sum_{n=1}^{k} a_n \tag{1.5.29}$$

Therefore, so long as the series diverges, we have that $q_k q_{k-1} \to \infty$ as $k \to \infty$. □

1.6 Advantages and Disadvantages of Continued Fractions

The power and utility of continued fractions is best explained by Khintchine [Ki], and I will summarize some of his key points. First, we can represent every number $\alpha \in (0, 1)$ as a continued fraction, and we can compute α's value to any desired precision; this property is especially useful in the case $\alpha \notin \mathbb{Q}$. In the next chapter, we will develop upper and lower bounds on a continued fraction's ability to approximate α, given the expansion has length N.

Many properties of α are revealed in its continued fraction expansion. For example, if a number is irrational, then the continued fraction never terminates. Moreover, Khintchine [Ki] comments: "Whilst every systematic fraction is coupled to a definite radix system (i.e. the base of a number system) and therefore unavoidably reflects more the interaction of the radix system and the number than the absolute properties of the number itself. The continued fraction is completely independent of any radix system and reproduces in pure form the properties of the number which it represents." Its independence of a radix system and the property that we can compute α's value to any desired precision make continued fractions extremely practical in both theoretical and practical settings. In fact, we will see shortly that a continued fraction gives the best possible rational approximation of an arbitrary $\alpha \in (0, 1)$.

There does exist one major disadvantage that was partially mentioned at the beginning of this section. Large expansions require considerable computational capital relative to decimal expansions. Furthermore, continued fractions do not lend themselves easily to arithmetical operations. For instance, the difficulty in adding, subtracting, multiplying, and dividing two or more continued fractions is often prohibitive, especially for large expansions.

Therefore, as a theoretical construct, continued fractions are extremely important, and computations involving a single continued fraction are usually cheap. But even basic manipulations of multiple continued fractions are too difficult to justify their use in such applications.

Chapter 2

Convergence and Approximation

2.1 Discussion and Bounds

The power of a continued fraction lies in its convergents ability to provide the best possible rational approximation to an irrational number α given denominator q. Assuming that $x = [a_0; a_1, a_2, \ldots]$ converges to a value α, the best rational approximations are provided by evaluating the convergents $\frac{p_n}{q_n}$ of x. By understanding how quickly the convergents of x converge to α, we can give exact answers to many questions involving measure and continued fractions. For example, what is the measure of the interval of uncertainty given a continued fraction expansion of length N? The answer to this question was given by Lemma 1.4.6, and the measure of the interval of uncertainty is $\mu\Big([\frac{p_{n-1}}{q_{n-1}}, \frac{p_n}{q_n}]\Big)$, where n is assumed to be odd and μ denotes the Lebesgue measure. Hereinafter, I will assume the reader has a working knowledge of Lebesgue measure theory.

In chapter 4, we try to explain and classify the behavior of continued fraction expansions with a large valued digit (i.e. $a_i = k_i$, where $k_i \gg 0$). A large coefficient value k_i directly affects the convergence rate of $x = [a_0; a_1, \ldots]$ to its value α; therefore, only through a proper understanding of continued fraction convergence will we understand the implications of observing a large coefficient value k_i. With our knowledge of convergence, we can address questions such as: do continued fractions converge to α faster than Kuzmin predicts? Finally, an understanding of convergence is necessary in order to follow the proof of Kuzmin's Theorem.

Our first theorem will present upper and lower bounds for $\left|\alpha - \frac{p_k}{q_k}\right|$.

Theorem 2.1.1. *For all $k \geq 0$ and for an irrational number α, we have:*

$$\frac{1}{q_k(q_k + q_{k+1})} < \left|\alpha - \frac{p_k}{q_k}\right| < \frac{1}{q_k q_{k+1}} \tag{2.1.1}$$

Proof: First we derive the lower bound. A consequence of Definition 1.2.6 is the intermediate fraction $\frac{p_k+p_{k+1}}{q_k+q_{k+1}}$ is enclosed between $\frac{p_k}{q_k}$ and α (see [Ki]). Therefore:

$$\left|\alpha - \frac{p_k}{q_k}\right| > \left|\frac{p_k + p_{k+1}}{q_k + q_{k+1}} - \frac{p_k}{q_k}\right| = \frac{1}{q_k(q_k + q_{k+1})}. \tag{2.1.2}$$

An equality sign is not possible because then $\alpha = \frac{p_k+p_{k+1}}{q_k+q_{k+1}} = \frac{p_{k+2}}{q_{k+2}}$, which implies that $a_{k+2} = 1$; in this case, α would have a terminating continued fraction expansion, but we assumed α to be irrational.

Now we develop the upper bound by following the exposition in [MT]. Consider the continued fraction $\alpha = [a_0; a_1, \ldots, a_n, a_{n+1}, \ldots] = [a_0; a_1, \ldots, a_n, a'_{n+1}]$, where $a'_{n+1} > a_{n+1}$ and is irrational. So we write:

$$\left|\alpha - \frac{p_n}{q_n}\right| = \left|\frac{a'_{n+1}p_n + p_{n-1}}{a'_{n+1}q_n + q_{n-1}} - \frac{p_n}{q_n}\right| = \left|\frac{(-1)^n}{q_n q'_{n+1}}\right|, \tag{2.1.3}$$

whence, $q'_{n+1} = a'_{n+1}q_n + q_{n-1} > a_{n+1}q_n + q_{n-1} = q_{n+1}$. Therefore:

$$\begin{aligned} \left|\alpha - \frac{p_n}{q_n}\right| &= \left|\frac{(-1)^n}{q_n q'_{n+1}}\right| \\ but\ q'_{n+1} &> q_{n+1} \\ so\ \left|\alpha - \frac{p_n}{q_n}\right| &< \frac{1}{q_n q_{n+1}}. \quad \square \end{aligned} \tag{2.1.4}$$

2.2 Convergents as Best Approximations

Definition 2.2.1. *The rational number $\frac{a}{b}$ ($b > 0$) is a best approximation to $\alpha \in (0, 1)$ if every other rational fraction having the same or smaller denominator differs from α more than $\frac{a}{b}$, or formally for $0 < d \leq b$, and $\frac{a}{b} \neq \frac{c}{d}$, we have:*

$$\left|\alpha - \frac{c}{d}\right| > \left|\alpha - \frac{a}{b}\right|. \tag{2.2.5}$$

We call $\frac{a}{b}$ a best approximation of the first kind. [Ki]

Definition 2.2.2. *The rational number $\frac{a}{b}$, $(b > 0)$ is a best approximation of the second kind to $\alpha \in (0,1)$ if for $0 < d \leq b$ and for $\frac{a}{b} \neq \frac{c}{d}$, we have: [Ki]*

$$|d\alpha - c| > |b\alpha - a|. \tag{2.2.6}$$

Every best approximation of the second kind is also a best approximation of the first kind, but the converse does not hold. To see that the converse does not hold, consider that $\frac{1}{3}$ is a best approximation of the first kind to $\frac{1}{5}$, but $\frac{0}{1}$ is the best approximation of the second kind.

We will merely state the following theorems, all of which justify our exclusive focus on convergents as the means, by which unbounded coefficients affect a continued fraction's value x.

Theorem 2.2.3. *Every best approximation of the first kind to the number α is either a convergent or an intermediate fraction of the continued fraction which represents this number. [Ki]*

Theorem 2.2.4. *Every best approximation of the second kind is a convergent. [Ki]*

The converse of the previous theorem is also true:

Theorem 2.2.5. *Every convergent is a best approximation of the second kind. The only trivial exception is given by: [Ki]*

$$\alpha = a_0 + \frac{1}{2}. \tag{2.2.7}$$

2.3 Absolute Difference Approximation

We will now attempt to refine our estimate of the difference $\left|\alpha - \frac{p_n}{q_n}\right|$. Recall from Theorem 2.1.1 that $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2}$, where we use the fact $q_n < q_{n+1}$. This section will answer if we can replace the right hand side of this inequality by some other function of q_n, which would result in a smaller upper bound than $\frac{1}{q_n^2}$. Whatever refinement we consider, it must apply to all α and hold for most n. As Khintchine [Ki] described the problem, how small can ϵ be such that we cannot find an $\alpha \in (0,1)$ satisfying $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1-\epsilon}{q_n^2}$ for only a finite number of n. In other words, the inequality $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1-\epsilon}{q_k^2}$ must hold without exception for any α and for infinitely many n.

Theorem 2.3.1. *If the number α possesses a convergent of order $n > 0$, then at least one of the two inequalities:*

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{2q_n^2}$$
$$\left|\alpha - \frac{p_{n-1}}{q_{n-1}}\right| < \frac{1}{2q_{n-1}^2} \qquad (2.3.8)$$

holds. *[Ki]*

Proof: We showed as a consequence of Theorem 1.4.7 that $\alpha \in \left[\frac{p_{n-1}}{q_{n-1}}, \frac{p_n}{q_n}\right]$ for odd n, but when n is even, the endpoints of the interval are switched. Now consider the sum:

$$\left|\alpha - \frac{p_n}{q_n}\right| + \left|\alpha - \frac{p_{n-1}}{q_{n-1}}\right| = \left|\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}}\right| = \frac{1}{q_n q_{n-1}} < \frac{1}{2q_n^2} + \frac{1}{2q_{n-1}^2} \qquad (2.3.9)$$

The last inequality holds because the geometric mean of $\frac{1}{q_n^2}$ and $\frac{1}{q_{n-1}^2}$ (i.e.$\frac{1}{q_n q_{n-1}}$) is less than their arithmetic mean (i.e. $\frac{1}{2q_n^2} + \frac{1}{2q_{n-1}^2}$). Equality would be possible only if $q_n = q_{n-1}$. $\square$

The following theorem is in a sense the converse of the theorem just proved.

Theorem 2.3.2. *Every irreducible fraction $\frac{a}{b}$ satisfying the inequality $\left|\alpha - \frac{a}{b}\right| < \frac{1}{2b^2}$ is a convergent of the number α.* *[Ki]*

Proof: I will not provide the proof because we do not need this theorem's results, but the proof is found in [Ki].

In Theorem 2.3.1, the smallest constant we could find was $\epsilon = \frac{1}{2}$, which is better than our starting bound by a factor of one-half, but this choice of ϵ does not hold for all n. The next two theorems along with Liouville's Theorem will provide the "supremum" for the difference $\left|\alpha - \frac{p_n}{q_n}\right|$.

The following theorem is very important, but the proof will not shed much light on our discussion so I refer the reader to [Ki].

Theorem 2.3.3. *If α possesses a convergent of order $n > 1$, then at least one of the three inequalities below will be satisfied:*

$$\begin{aligned}\left|\alpha - \frac{p_n}{q_n}\right| &< \frac{1}{\sqrt{5}q_n^2}\\ \left|\alpha - \frac{p_{n-1}}{q_{n-1}}\right| &< \frac{1}{\sqrt{5}q_{n-1}^2}\\ \left|\alpha - \frac{p_{n-2}}{q_{n-2}}\right| &< \frac{1}{\sqrt{5}q_{n-2}^2}. \end{aligned} \qquad (2.3.10)$$

The results of this theorem are profound because they will be used to prove $\epsilon = \frac{1}{\sqrt{5}}$ is the smallest constant, such that $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{\epsilon}{q_n^2}$ holds for all α and for an infinite number of n (but not all n). This claim is valid because for any constant c smaller than $\epsilon = \frac{1}{\sqrt{5}}$, we can find an α such that $\left|\alpha - \frac{p_n}{q_n}\right| > \frac{c}{q_n^2}$, namely $\alpha = \frac{1+\sqrt{5}}{2}$ or $\alpha = [1; 1, 1, 1, \ldots]$. For this α we have $\left|\alpha - \frac{p_n}{q_n}\right| = \frac{1}{q_n^2(\sqrt{5}+\epsilon_n')}$, where ϵ_n' is the uncertainty in any approximation $\frac{p_n}{q_n}$ of $\alpha \in \mathbb{R}$, and $\epsilon \to 0$ as $n \to \infty$. Thus, if $c < \frac{1}{\sqrt{5}}$, then $\left|\alpha - \frac{p_n}{q_n}\right| > \frac{c}{q_n^2}$ for $\alpha = \frac{1+\sqrt{5}}{2}$.

Theorem 2.3.4. *For any $\alpha \in \mathbb{R}$ the inequality $\left|\alpha - \frac{p}{q}\right| < \frac{c}{q^2}$ has infinitely many solutions $p, q \in \mathbb{Z}$ $(q > 0)$ if $c \geq \frac{1}{\sqrt{5}}$. On the other hand, if $c < \frac{1}{\sqrt{5}}$, then it is possible to find an α, such that $\left|\alpha - \frac{p}{q}\right| < \frac{c}{q^2}$ has no more than a finite number of solutions. [Ki]*

Proof: This proof is taken from Khintchine's exposition [Ki]. We showed the second statement of the theorem by noting for any $c < \frac{1}{\sqrt{5}}$ we can take $\alpha = [1; 1, 1, 1, 1, \ldots] = \frac{1+\sqrt{5}}{2}$, which satisfies $\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^2}$. The proof of the first statement relies on Theorem 2.3.3. Let α be irrational, which implies an infinite continued fraction expansion. From Theorem 2.3.3, we know $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{\sqrt{5}q_n^2}$ is satisfied at least once in every three convergents for every $\alpha \in \mathbb{R}$, but since $\alpha \notin \mathbb{Q}$, this inequality is satisfied infinitely often. Thus the first statement is proved for irrational α. Now let $\alpha \in \mathbb{Q}$, then α can be represented as $\alpha = \frac{c}{d}$. Assuming the expansion of α at least has digits a_1, a_2, a_3, then this expansion possesses at least three convergents. From Theorem 2.3.3, we know at least one in every three convergents satisfies $|\alpha - \frac{p}{q}| < \frac{c}{q^2}$, so let $\frac{p}{q}$ be the convergent satisfying this inequality and take $q = nd, p = nc, (n = 1, 2, \ldots)$. □

The previous theorem states rigorously that we cannot find an approximating constant c smaller than $\frac{1}{\sqrt{5}}$ such that

$$\left|\alpha - \frac{p}{q}\right| < \frac{c}{q^2} \tag{2.3.11}$$

holds for all α and all q. However, we will see in the next theorem that no matter how small c may be, we can always find an $\alpha \in \mathbb{R}$, such that the inequality in Equation 2.3.11 is satisfied.

Theorem 2.3.5. *Define a function of q such that for all q, $f(q) > 0$. Then regardless of the behavior of $f(q)$, we can find an irrational α, such that the inequality:*

$$\left|\alpha - \frac{p}{q}\right| < f(q) \tag{2.3.12}$$

should possess infinitely many solutions $p, q \in \mathbb{Z},\ (q > 0)$.

Proof: By controling the behavior of the coefficients of a continued fraction expansion, we can construct an α, such that the claim of the theorem is satisfied. We assume α is irrational, and we impose the following restriction on a_n for all $n = (1, 2, \ldots)$:

$$a_{n+1} > \frac{1}{q_n^2 f(q_n)}. \tag{2.3.13}$$

Note that we can find infinitely many α satisfying Inequality 2.3.13. Therefore, for any $n \geq 0$, we have the following:

$$\begin{aligned} \left|\alpha - \frac{p_n}{q_n}\right| &< \frac{1}{q_n q_{n+1}} = \frac{1}{q_n(a_{n+1}q_n + q_{n-1})} \\ &\leq \frac{1}{a_{n+1}q_n^2} < f(q_n), \end{aligned} \tag{2.3.14}$$

where the last step is justified by Inequality 2.3.13. □

The next argument will help us understand the effects of a large valued coefficient on x's convergence to α (see Chapter 4). For α irrational, we proved in

Theorem 2.1.1:

$$\begin{aligned} \frac{1}{q_n(q_n+q_{n+1})} < &\left|\alpha - \tfrac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} \\ \Rightarrow \frac{1}{q_n^2(a_{n+1}+1+\frac{q_{n-1}}{q_n})} < &\left|\alpha - \tfrac{p_n}{q_n}\right| < \frac{1}{q_n^2(a_{n+1}+\frac{q_{n-1}}{q_n})} \\ \Rightarrow \frac{1}{q_n^2(a_{n+1}+2)} < &\left|\alpha - \tfrac{p_n}{q_n}\right| < \frac{1}{q_n^2 a_{n+1}}, \end{aligned} \tag{2.3.15}$$

where the left hand side of the second line is a result of expanding line 1, factoring q_n out of the denominator, and then utilizing the definition of $\frac{q_{n+1}}{q_n}$ from Equation 1.4.10. The right hand side of line 2 is a consequence of letting $q_{n+1} = q_{n-1} + q_n a_{n+1}$, expanding, and then factoring q_n^2 out of the denominator. The last step is basic algebra.

If $a_{n+1} = k_{n+1}$ for k_{n+1} arbitrarily large, then the convergent $\frac{p_n}{q_n}$ becomes very close to α in Equation 2.3.15. This large value of k_{n+1} will increase the precision of x's approximation of α because all convergents subsequent to $\frac{p_n}{q_n}$ are closer to α than $\frac{p_n}{q_n}$ is (see Theorem 1.4.7). In other words, the result of observing a large coefficient value k_{n+1} is a large reduction in the measure of the interval of uncertainty.

Extending our analysis, irrationals, whose continued fraction expansions are characterized by the frequent occurrence of large valued digits, are approximated well by rationals because the convergents $\frac{p_n}{q_n}$ converge to α very quickly. However, it may be the case that α, whose expansion has only a few extremely large valued coefficients and also is replete with small valued coefficients, is not approximated by rationals better than $\alpha^{'}$, whose expansion has no extremely large valued coefficients and far fewer small valued coefficients. The worst approximated irrational is obviously $\alpha = [1; 1, 1, \ldots]$, which explains why this irrational is always the limiting case for making generalized statments about continued fraction convergence.

The next theorem extends our previous results, in that irrationals with bounded coefficients cannot be approximated to a degree better than $\frac{1}{q^2}$; however, the set of such α has zero measure (see Chapter 4). On the other hand, irrationals with unbounded coefficients can be approximated to a degree far better than $\frac{1}{q^2}$

Theorem 2.3.6. *For every irrational number α with bounded elements, the inequality:*

$$\left|\alpha - \frac{p}{q}\right| < \frac{c}{q^2} \tag{2.3.16}$$

has for sufficiently small c no solution in $p, q \in \mathbb{Z}$ $(q > 0)$. Conversely, for every α with unbounded elements, the above inequality is satisfied for arbitrary $c > 0$, by infinitely many integers p, q. [Ki]

Proof: The proof of this theorem is found in [Ki]. First, we will prove the second assertion. Let $c > 0$ be arbitrary. Since the continued fraction expansion of an irrational α is infinite and we assume the coefficients are unbounded, then we can find an infinite number of values of n such that $a_{n+1} > \frac{1}{c}$. Thus, from the last inequality in Equation 2.3.15, we have for infinitely many n:

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{c}{q_n^2}. \tag{2.3.17}$$

Now, we prove the first assertion by finding a sufficiently small c, such that there exists no $p, q \in \mathbb{Z}$ satisfying Equation 2.3.16 for an irrational α with bounded coefficients. Because α has bounded coefficients, there exist $M > 0$, such that $a_n < M$ for all $n \in \mathbb{N}$. Substituting M for a_n in Equation 2.3.15, we observe:

$$\frac{1}{q_n^2(M+2)} < \left|\alpha - \frac{p_n}{q_n}\right| \tag{2.3.18}$$

for any $n \geq 0$. Now, let p, q be arbitrary integers and set the index n according to the condition $q_{n-1} < q < q_n$. By Theorem 2.2.5 and by the fact that all best approximations of the second kind are also best approximations of the first kind, we conclude all convergents are best approximations of the first kind. Therefore, $\left|\alpha - \frac{p_n}{q_n}\right| \leq \left|\alpha - \frac{p}{q}\right|$ for all $q \leq q_n$, and we argue:

$$\begin{aligned}
\left|\alpha - \frac{p}{q}\right| &\geq \left|\alpha - \frac{p_n}{q_n}\right| > \frac{1}{q_n^2(M+2)} \\
&= \frac{1}{q^2(M+2)}\left(\frac{q}{q_n}\right)^2 > \frac{1}{q^2(M+2)}\left(\frac{q_{n-1}}{q_n}\right)^2 \\
&= \frac{1}{q^2(M+2)}\left(\frac{q_{n-1}}{a_n q_{n-1} + q_{n-2}}\right)^2 \\
&> \frac{1}{q^2(M+2)}\frac{1}{(a_n+1)^2} > \frac{1}{(M+2)(M+1)^2 q^2} \\
\Rightarrow \text{choose} \quad c &< \frac{1}{(M+2)(M+1)^2}.
\end{aligned} \tag{2.3.19}$$

The first equation is justified by Theorem 2.2.5 and Equation 2.3.18. The inequality in the second line is justified by our construction of the index n. The third step is simply expanding q_n according to Theorem 1.4.2. The fourth line is obtained by multiplying the numerator and the denominator of the second factor in line 3 by $\frac{1}{q_{n-1}} \Rightarrow \frac{1}{a_n + \frac{q_{n-2}}{q_{n-1}}}$ and by recognizing Theorem 1.4.2 gives $\frac{q_{n-2}}{q_{n-1}} < 1$. To arrive at the last inequality in line 4, we note $a_n < M$. Thus, for c satisfying the final inequality, there does not exist $p, q \in \mathbb{Z}$, such that $\left|\alpha - \frac{p}{q}\right| < \frac{c}{q^2}$ is satisfied. □

We found $c = \frac{1}{\sqrt{5}}$ is the smallest constant such that Equation 2.3.11 holds for all α and for most n, but can we improve the degree of q_n^2 in this same equation (i.e. can we find a bound in Equation 2.3.11 of the form $\frac{c}{q^{2+\epsilon}}$)? The answer happens to be no.

Theorem 2.3.7. *Let C, ϵ be positive constants. Let S be the set of all points $\alpha \in [0, 1]$, such that there exist infinitely many coprime integers p, q satisfying:*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{C}{q^{2+\epsilon}}. \tag{2.3.20}$$

Then the measure of S is zero, denoted by $\mu(S) = 0$. [MT]

Proof: This proof is an excerpt from [MT]. Let $N > 0$ and define S_N to be the set consisting of all the $\alpha \in [0, 1]$, such that there exists $p, q \in \mathbb{Z}$ and $q > N$, for which:

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{C}{q^{2+\epsilon}}. \tag{2.3.21}$$

Since we defined S to be the set of α for which there exists an infinite number of coprime pairs of p, q satisfying Inequality 2.3.21, we have if $\alpha \in S$ then $\alpha \in S_N$ for all N. Otherwise, α has only a finite number N of denominators, for which the inequality can be made to hold, and because there are at most $q + 1$ choices of p for each denominator q, the maximum number of coprime pairs of p, q satisfying Inequality 2.3.21 is finite. Therefore, it suffices to show that as $N \to \infty$, $\mu(S_N) \to 0$.

We must estimate the size of S_N. Let $\frac{p}{q}$ be given, and consider the measure of the set of all α within $\frac{C}{q^{2+\epsilon}}$ of $\frac{p}{q}$. This set is an interval:

$$I_{p,q} = \Big(\frac{p}{q} - \frac{C}{q^{2+\epsilon}}, \frac{p}{q} + \frac{C}{q^{2+\epsilon}}\Big), \tag{2.3.22}$$

and the measure of this interval is $\frac{2C}{q^{2+\epsilon}}$. Now, let I_q be the set of all $\alpha \in [0,1]$ within $\frac{C}{q^{2+\epsilon}}$ of all rational numbers with denominator q; this set will be the sum of at most q intervals. Then we have:

$$\begin{aligned} I_q &\subseteq \bigcup_{p=0}^{q} I_{p,q} \\ \Rightarrow \mu(I_q) &\leq \sum_{p=0}^{q} \mu(I_{p,q}) = (q+1)\frac{2C}{q^{2+\epsilon}} \\ &= \frac{q+1}{q}\frac{2C}{q^{1+\epsilon}} \leq \frac{4C}{q^{1+\epsilon}} \end{aligned} \tag{2.3.23}$$

where the last inequality holds because $q \in \mathbb{Z}$ and $1 \leq q$. Now we can consider $\mu(S_N)$:

$$\begin{aligned} \mu(S_N) &\leq \sum_{q>N}^{\infty} \mu(I_q) \\ &\leq \sum_{q>N}^{\infty} \frac{4C}{q^{1+\epsilon}} \\ &< \frac{4C}{\epsilon}N^{-\epsilon} \end{aligned} \tag{2.3.24}$$

We know that the series converges because $\epsilon > 0$. The first step is justified by the definitions of S_N and I_q. Applying Equation 2.3.23 to line 1 yields line 2. The last step is justified by recognizing $\sum_{q>N}^{\infty} \frac{1}{q^{1+\epsilon}} < \int_{q=N}^{\infty} \frac{1}{q^{1+\epsilon}} = \frac{1}{\epsilon}N^{-\epsilon}$. Thus as $N \to \infty$, $\mu(S_N) \to 0 \Rightarrow \mu(S) \to 0$ because $S \subset S_N$. □

The previous theorem implies that except for a set of zero measure, one cannot find a rational number $\frac{p}{q}$ that approximates α better than $\frac{C}{q^2}$. See Theorem 2.4.1 for a more general result.

2.4 Liouville's Theorem

In this section we present Lioville's Theorem, which allows us to define families for the numbers examined in Chapters 3 and 4. In Chapter 3 we examine degree 3 algebraic irrational numbers, and in Chapter 4 we analyze algebraic numbers of degrees 3,5,7,11,13. Additionally, Liouville's Theorem will provide one final approximation of the difference $\left|\alpha - \frac{p}{q}\right|$.

Theorem 2.4.1. *Corresponding to every algebraic number of degree n, there exists a constant $C > 0$, such that for any integers p, q $(q > 0)$, we have:*

$$\left|\alpha - \frac{p}{q}\right| > \frac{C}{q^n}. \quad [Ki] \tag{2.4.25}$$

Proof: The proof is not included because it does not add any additional intuition or insight into the properties of continued fractions, on which I am focusing. See [Ki] or [MT] for the proof.

This theorem implies an algebraic number α cannot be approximated by a rational fraction to a degree of accuracy exceeding α's algebraic degree. An interesting result of Liouville's Theorem is a method for constructing transcendental numbers.

A summary of the method is as follows: let $C > 0$ be arbitrary and let m be any positive integer. If α satisfies

$$\left|\alpha - \frac{p}{q}\right| < \frac{C}{q^m} \tag{2.4.26}$$

for some p, q, then α is transcendental. After the coefficients $a_1, a_2, \ldots, a_n$ have been chosen, take $a_{n+1} > q_n^{n-1}$. Then we have:

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2 a_{n+1}} < \frac{1}{q_n^{n+1}}, \tag{2.4.27}$$

where the first inequality is a result of Theorem 2.1.1. The second inequality is a result of Theorem 1.4.2, and the third inequality is from the condition $a_{n+1} > q_n^{n-1}$. Thus, for sufficiently large values of a_i, Inequality 2.4.26 is satisfied for arbitrary C and m. Now, we can reason if α's continued fraction expansion possesses 'too many' extremely large valued coefficients, then α may be transcendental.

Chapter 3

Kuzmin's Theorem and Levy's Improved Bound

This chapter will provide the crux of the theory needed for this thesis. I will present some necessary definitions and lemmas, Kuzmin's Theorem, and then Levy's refinement of Kuzmin's Theorem. However, the main focus of this chapter will be to trace the derivation of not only the closed form expression for $Prob(a_n = k_n)$, where a_n is a function of α but we will denote $a_n(\alpha)$ simply as a_n, but also the error estimate for this expression. As we derive this expression, we will obtain some results that can explain our empirical observations, which are presented in this chapter and in Chapter 4.

3.1 The Gaussian Problem

Let $l(\alpha, N)$ denote the length of the continued fraction expansion of the number α expanded to N coefficients. Because the expansion of $\alpha \in \mathbb{Q}$ is finite, we can assume α is irrational without any loss of generality. In a letter dated January 30, 1812, to Laplace, Gauss mentioned that he was unable to find a closed form solution to the following problem: let $M \in [0, 1]$ be an unknown, for which all values are either equally possible or given according to some distribution law. Then convert M into its simple continued fraction form:

$$M = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}. \tag{3.1.1}$$

We ask what the probability is when stopping the expansion at a finite term a_N that the tail

$$M^{'} = \cfrac{1}{a_{n+1} + \cfrac{1}{a_{n+2} + \cfrac{1}{\ddots}}} \tag{3.1.2}$$

will represent a number $B \in (0, x)$, where $x \leq 1$? We label this probability $P(N, x)$, and if all values of M are equally probable, then $P(0, x) = x$. Gauss was able to prove as $N \to \infty$:

$$\lim_{N \to \infty} P(N, x) = \frac{\log(1 + x)}{\log 2}, \tag{3.1.3}$$

but for large N, Gauss wanted to find an explicit estimate for the difference:

$$P(N, x) - \frac{\log(1 + x)}{\log 2} \tag{3.1.4}$$

In 1928, Kuzmin was the first person to estimate this difference. In his proof, Kuzmin assumed M is uniformly distributed on the interval $[0, 1]$. In 1929, Levy used an entirely different approach to estimate the difference in Equation 3.1.4, and he gave a better estimate of the difference than Kuzmin did.

3.2 Intervals of Rank n

3.2.1 Definition and Intuition

Recall we used the Euclidean algorithm to find the coefficients of a continued fraction expansion. This method yielded $a_1 = [\frac{1}{\alpha}]$, where $[x]$ is the smallest integer not greater than x. Therefore, we note $a_1 = 1$ for $1 \leq \frac{1}{\alpha} < 2 \Rightarrow \frac{1}{2} < \alpha \leq 1$, where the strict inequality is due to the greatest integer function $[\ ]$. We can proceed in this fashion, such that we arrive at the general case:

$$a_1 = k \ \ for \ \ k \leq \frac{1}{\alpha} < k+1 \Rightarrow \frac{1}{k+1} < \alpha \leq \frac{1}{k} \tag{3.2.5}$$

A notable property about a_1 is it assumes a constant value on the interval $(\frac{1}{k+1}, \frac{1}{k}]$, and if $a_1 = k$, then $\alpha \in (\frac{1}{k+1}, \frac{1}{k}]$. Secondly, a_1 is discontinuous at every integer. Lastly, the area under the a_1 function is given by:

$$\int_0^1 a_1(\alpha)\, d\alpha = \infty. \tag{3.2.6}$$

This result follows from recognizing each interval $(\frac{1}{k+1}, \frac{1}{k}]$ is a rectangle of width $\frac{1}{k} - \frac{1}{k+1}$ and height k, so writing the divergent integral as an infinite series yields:

$$\int_0^1 a_1(\alpha)\, d\alpha = \infty = \sum_{k=1}^{\infty} k\Big(\frac{1}{k} - \frac{1}{k+1}\Big) = \sum_{k=1}^{\infty} \frac{1}{k+1}. \tag{3.2.7}$$

Definition 3.2.1. *Let $a_1 = k$, where k is an interger, then $(\frac{1}{k+1}, \frac{1}{k}]$ is an interval of rank one, which implies $\alpha \in (\frac{1}{k+1}, \frac{1}{k}]$.*

We can perform the same procedure for a_2 by fixing $a_1 = k$ and taking $a_2 = [r_2]$, where r_2 can take any value in $[1, \infty)$. Letting $a_1 = k$ and $a_2 = h$, the interval of rank two corresponding to numbers whose first two digits are k and h is given by:

$$\Big(\frac{1}{k+\frac{1}{h}}, \frac{1}{k+\frac{1}{h+1}}\Big). \tag{3.2.8}$$

Assuming $a_1 = k$ is given, the intuition of the equation above is that after the value of a_2 is ascertained, we have "shortened" the interval, in which α can reside $\Big($i.e. $\alpha \in \Big(\frac{1}{k+\frac{1}{h}}, \frac{1}{k+\frac{1}{h+1}}\Big)\Big)$.

Comparing the interval of rank one with the interval of rank two, as $k \to \infty$ in the interval of rank one, we have $\alpha \to 0$. Thus, as $k \to \infty$, the rank one intervals form a sequence proceeding from right to left and are indexed by k; this is the case for all intervals of rank n, where n is odd. On the other hand, let $a_1 = k$ be fixed, then as $h \to \infty$ in the intervals of rank two, we have $\alpha \to \frac{1}{k}$, or the larger endpoint of a rank two interval. This results implies that for n even, the intervals of rank n form a sequence that runs from left to right. The formal expression giving the directions of these sequences is provided by considering:

$$\alpha - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(q_n r_{n+1} + q_{n-1})} \tag{3.2.9}$$

We now define an interval of rank n.

Definition 3.2.2. *Let $a_1 = k_1, a_2 = k_2, \ldots, a_n = k_n$ be given, then the corresponding interval of rank n is*

$$J_n = \Big(\frac{p_n}{q_n}, \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\Big]. \tag{3.2.10}$$

This interval is the formal expression of the previously defined ***interval of uncertainty*** *given a continued fraction expansion of length n.*

The endpoints of a rank n interval are easily obtained. First, we recognize $\alpha = [k_1, \ldots, k_n, r_{n+1}] \Rightarrow \alpha = \frac{p_n r_{n+1} + p_{n-1}}{q_n r_{n+1} + q_{n-1}}$. Let $1 \leq r_{n+1} < \infty$. Substituting in $r_{n+1} = 1$ gives $\alpha = \frac{p_n + p_{n-1}}{q_n + q_{n-1}}$, which is the larger endpoint of J_n. Now let $r_{n+1} \to \infty$, and we obtain:

$$\lim_{r_{n+1} \to \infty} \frac{p_n r_{n+1} + p_{n-1}}{q_n r_{n+1} + q_{n-1}} = \frac{p_n}{q_n}, \tag{3.2.11}$$

which is the smaller endpoint of J_n. Additionally, as r_{n+1} runs through the natural numbers, the interval J_n is partitioned into a countable number of intervals of rank $n+1$; the sequence of rank $n+1$ intervals runs from left to right for n odd, and right to left for n even. Also, α is a monotonic function of r_{n+1} for $r_{n+1} \in [1, \infty)$.

The next argument is extremely useful in forthcoming proofs, and the logic is adopted from [MT].

Consider all continued fraction expansions of the form $\alpha = [0; a_1, \ldots, a_n]$ and look at all n-tuples (i.e. *all* combinations of *all* values k_i for *every* a_i). Given a rank n interval J_n defined by the continued fraction expansion beginning with $[0; a_1, \ldots, a_n]$, we can find the enpoints of J_n by taking the infinite union over all rank $n-1$ subintervals comprising J_n:

$$J_n = \bigcup_{k=1}^{\infty} \Big(\frac{p_n(k+1) + p_{n-1}}{q_n(k+1) + q_{n-1}}, \frac{p_n(k) + p_{n-1}}{q_n(k) + q_{n-1}}\Big] = \Big(\frac{p_n}{q_n}, \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\Big], \tag{3.2.12}$$

where we let $a_{n+1} = k$ for $k = 0, 1, 2, \ldots$, and the endpoints of J_n are given by letting both $k = 1$ and $k \to \infty$. The length of each rank n interval is $\Big|\frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\Big|$.

Each $\frac{p_n}{q_n}$ corresponds to a unique, disjoint rank n subinterval of $[0,1]$, as each possible continued fraction expansion of length n leads to a different interval of rank n; this is easily seen by applying Theorem 1.4.2 n times to Definition 3.2.2. From [MT], we conclude:

$$\begin{aligned} [0,1] &= \bigcup_{(a_1,\dots,a_n)\in N^n} \left(\frac{p_n}{q_n}, \frac{p_n+p_{n-1}}{q_n+q_{n-1}}\right] \\ 1 &= \sum_{(a_1,\dots,a_n)\in N^n} \left|\frac{p_n}{q_n} - \frac{p_n+p_{n-1}}{q_n+q_{n-1}}\right|. \end{aligned} \tag{3.2.13}$$

Therefore, as argued previously, the union of all rank n intervals covers $[0,1]$.

3.2.2 Prob($a_n = k$)

In this section we examine the set of points $\alpha \in [0,1]$, such that $a_n = k$. From our analysis of Equation 3.2.12, we expect this set to be a union of rank n intervals. The measure of this union will be equal to the measure of the set of $\alpha \in [0,1]$, such that $a_n = k$. What is the length or measure μ of the union of these intervals?

We denote by $E\begin{pmatrix} 1 & 2\cdots n \\ k_1 & k_2\cdots k_n \end{pmatrix}$ the set of $\alpha \in [0,1]$, such that $a_1 = k_1,\ a_2 = k_2,\dots,\ a_n = k_n$, which clearly defines an interval of rank n. If we let the values of k_i be arbitrary, then $E\begin{pmatrix} 1 & 2\cdots n \\ k_1 & k_2\cdots k_n \end{pmatrix}$ defines an arbitrary rank n interval, J_n; correspondingly we let $J_{n+1}^s = E\begin{pmatrix} 1 & 2\cdots n & n+1 \\ k_1 & k_2\cdots k_n & s \end{pmatrix}$ be an interval of rank $n+1$ contained in J_n (s is used in order to show the expression $Prob(a_{n+1} = k_{n+1})$'s independence from n and from k_i for $i \le n$).

Recall that $\alpha = \frac{p_n r_{n+1} + p_{n-1}}{q_n r_{n+1} + q_{n-1}}$, and if $a_{n+1} = [r_{n+1}] = s$, then $r_{n+1} \in [s, s+1)$. Therefore:

$$\alpha \in \left[\frac{p_n s + p_{n-1}}{q_n s + q_{n-1}}, \frac{p_n(s+1) + p_{n-1}}{q_n(s+1) + q_{n-1}}\right). \tag{3.2.14}$$

This interval denotes the endpoints of J_{n+1}^s, and we have $J_{n+1}^s \subset J_n$ for some J_n. Thus, we must estimate $\mu(J_{n+1}^s)$ by using a conditional probability argument, where μ denotes the Lebesgue measure, or in our case, the linear difference between the two endpoints.

From Definition 3.2.2 and Equation 3.2.14, we have:

$$\mu(J_n) = \left|\frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right| \quad = \quad \frac{1}{q_n^2(1 + \frac{q_{n-1}}{q_n})}$$

$$\mu(J_{n+1}^s) = \left|\frac{p_n s + p_{n-1}}{q_n s + q_{n-1}} - \frac{p_n(s+1) + p_{n-1}}{q_n(s+1) + q_{n-1}}\right| \quad = \quad \frac{1}{q_n^2 s^2(1 + \frac{q_{n-1}}{q_n})(1 + \frac{1}{s} + \frac{q_{n-1}}{sq_n})},$$

which implies

$$\frac{\mu(J_{n+1}^s)}{\mu(J_n)} = \frac{1}{s^2}\frac{1 + \frac{q_{n-1}}{q_n}}{(1 + \frac{q_{n-1}}{sq_n})(1 + \frac{1}{s} + \frac{q_{n-1}}{sq_n})}, \tag{3.2.15}$$

where μ denotes the Lebesgue measure. The right hand sides of the above equations are obtained by simple algebraic manipulations.

In order to bound the last expression from above in Equation 3.2.15, we need to make the numerator as large as possible and the denominator as small as possible. By Theorems 1.4.2 and 1.4.10, we have $\frac{q_{n-1}}{q_n} < 1$; so, let both $\frac{q_{n-1}}{q_n} \to 1$ and $s \to \infty$ in the second factor $\left(\frac{1+\frac{q_{n-1}}{q_n}}{(1+\frac{q_{n-1}}{sq_n})(1+\frac{1}{s}+\frac{q_{n-1}}{sq_n})}\right)$ on the right hand side of Equation 3.2.15, and we obtain the upper bound $\frac{1}{s^2}\frac{2}{(1+\epsilon)(1+\epsilon'+\epsilon)} = \frac{2}{s^2}$, where ϵ is an arbitrarily small constant. To establish the lower bound, we let both $\frac{q_{n-1}}{q_n} \to 1$ and $s = 1$ in the second factor, and then we obtain the lower bound $\frac{1}{s^2}\frac{2}{(2)(3)} = \frac{1}{3s^2}$. Collecting the results yields:

$$\frac{1}{3s^2} < \frac{\mu(J_{n+1}^s)}{\mu(J_n)} < \frac{2}{s^2}, \tag{3.2.16}$$

where the inequalities are strict because $\frac{q_{n-1}}{q_n} < 1$. The intuition behind the previous equation tells us in a rank n interval J_n, the $(n+1)$ rank interval determined by $a_{n+1} = s$ occupies roughly a $\frac{1}{s^2}$ part of J_n. Further, note the bounds are independent of n and k_i because the intervals of rank n cover $[0,1]$ by Equation 3.2.13. We can now write:

$$\frac{\mu(J_n)}{3s^2} < \mu(J_{n+1}^s) < \frac{2\mu(J_n)}{s^2} \tag{3.2.17}$$

but from Equation 3.2.13 we see

$$\sum_{n=1}^{\infty}\mu(J_n)=1 \qquad \text{and} \qquad \sum_{n=1}^{\infty}\mu(J_{n+1}^s)=\mu(E\binom{n+1}{s})$$

$$\Rightarrow \frac{1}{3s^2} < \quad \mu(E\binom{n+1}{s}) \quad < \frac{2}{s^2} \tag{3.2.18}$$

where the first step is justified by recognizing that $\mu(J_n) > 0$ for all n, thereby permitting us to multiply all terms by $\mu(J_n)$. The second line is justified by recognizing the union of all rank n intervals covers $[0, 1]$. The second equality in line 2 expresses the measure of the union of all rank $n + 1$ intervals characterized by $a_{n+1} = s$. The third line is a consequence of collecting the results in lines 1 and 2.

Khintchine [Ki] writes that the “the measure of the set of points for which a certain element has a given value s, always lies between $\frac{1}{3s^2}$ and $\frac{2}{s^2}$” or an interval with magnitude of order $\frac{1}{s^2}$. Another important implication of this result is the upper and lower bounds of the measure of J_{n+1}^s are not dependent on n or on k_i for $i \leq n$, which is a consequence of Equation 3.2.13.

3.3 Kuzmin’s Theorem

Kuzmin’s Theorem states for almost all irrational $\alpha \in [0, 1]$ and all $k \in Z$, the following inequality holds:

$$\left|\mu(E\binom{n}{k}) - \frac{\ln(1+\frac{1}{k(k+2)})}{\ln 2}\right| < \frac{A}{k(k+1)}e^{-\lambda\sqrt{n-1}}, \tag{3.3.19}$$

where A and λ are absolute positive constants.

In this section, I will present the proof of Kuzmin’s Theorem, which requires a bit of preliminary work. A very easy to follow proof of Kuzmin’s Theorem is provided in both [MT] and [De], both of which follow Khintchine’s exposition [Ki]. In addition, I will also present a summary of Levy’s results but will not provide a full proof of his approach; however, the proof can be found in his original paper [Le] or a good summary can be found in [De].

3.3.1 Notation and Definitions

Let us denote the following:

$$\alpha = [0; a_1, a_2, \ldots, a_n, \ldots],$$

$$r_n = r_n(\alpha) = [a_n; a_{n+1}, \ldots],$$

$$z_n = z_n(\alpha) = r_n - a_n = [0; a_{n+1}, a_{n+2}, \ldots] \Rightarrow z_n \in [0, 1),$$

$$m_n(x) = \mu\Big(\{\alpha \in [0,1] : z_n(\alpha) < x\}\Big).$$

Property 3.3.1. *Consider the sequence of positive functions* $m_i(x)$ *as defined above:*

$$m_0(x), m_1(x), \ldots, m_n(x), \ldots \tag{3.3.20}$$

then,

$$m_{n+1}(x) = \sum_{k=1}^{\infty} \left(m_n\Big(\frac{1}{k}\Big) - m_n\Big(\frac{1}{k+x}\Big)\right), \tag{3.3.21}$$

where we assume $0 \leq x \leq 1$ *and* $0 \leq n$*.*

Proof: $z_n = [0; a_{n+1}, \ldots] = r_n - a_n$, where $r_n = a_n + \frac{1}{r_{n+1}}$, but since $z_{n+1} = [0; a_{n+2}, \ldots] = r_{n+1} - a_{n+1} \Rightarrow r_{n+1} = a_{n+1} + z_{n+1}$. Hence the equation:

$$z_n = a_n + \frac{1}{r_{n+1}} - a_n = \frac{1}{a_{n+1} + z_{n+1}} \tag{3.3.22}$$

implies we must have for $z_{n+1} < x$:

$$\frac{1}{k+x} < z_n \leq \frac{1}{k}, \tag{3.3.23}$$

which follows by direct substitution into Equation 3.3.22. The measure of the set z_n satisfying Equation 3.3.23 is:

$$m_n\Big(\frac{1}{k}\Big) - m_n\Big(\frac{1}{k+x}\Big). \tag{3.3.24}$$

Thus, our recurrent relationship is shown. Note, a_n is specified in Equation 3.3.24, while a_n remains undetermined in Equation 3.3.21.

Differentiating Equation 3.3.21 term by term with respect to x gives:

$$m'_{n+1}(x) = \sum_{k=1}^{\infty} \frac{1}{(k+x)^2} m'_n\Big(\frac{1}{k+x}\Big). \tag{3.3.25}$$

We must address why we can differentiate this series term by term. The proof is by induction on n. Let $\alpha \in [0,1]$, then we have $z_0(\alpha) = \alpha \Rightarrow m_0(x) = x$, whereby $m_0^{'}(x) = 1$. Now assume that $m_n^{'}(x)$ is bounded and continuous for all $n < n+1$ and for all $x \in [0,1]$. Then the right hand side of Equation 3.3.25 is bounded and continuous by our inductive assumption, which implies $m_{n+1}^{'}$ is also bounded and continuous for all n; thus, the series on the right hand side of Equation 3.3.25 is bounded, continuous, and equal to $m_{n+1}^{'}(x)$ for all n. Equation 3.3.25 has been shown by induction.

With Equation 3.3.21 in mind, we search for a recurrent function with similar behavior. Gauss proposed the following function:

Lemma 3.3.2. *Let C be an arbitrary constant, then*

$$\phi(x) = C \ln(1+x) \tag{3.3.26}$$

satisfies

$$\phi(x) = \sum_{k=1}^{\infty} \left(\phi\Big(\frac{1}{k}\Big) - \phi\Big(\frac{1}{k+x}\Big) \right), \tag{3.3.27}$$

where k is a positive integer and $x \in [0,1]$.

Proof: Consider the following argument, noting both $\frac{d}{dx}\Big(\ln(1+x)\Big) = \frac{1}{1+x}$ and the first line in the following equation is a telescoping series:

$$\begin{aligned}
\frac{1}{1+t} &= \sum_{k=1}^{\infty} \frac{1}{(k+t)^2(1+\frac{1}{k+t})} \\
\Rightarrow \int_0^x \frac{1}{1+t}\,dt &= \sum_{k=1}^{\infty} \int_0^x \frac{1}{(k+t)^2(1+\frac{1}{k+t})}\,dt \\
\Rightarrow C\log(1+x) &= \sum_{k=1}^{\infty} \log\left(\frac{1+\frac{1}{k}}{1+\frac{1}{k+x}}\right) \\
\Rightarrow \log(1+x) &= \sum_{k=1}^{\infty} \left(\log\Big(1+\frac{1}{k}\Big) - \log\Big(1+\frac{1}{k+x}\Big) \right),
\end{aligned}$$

where we introduced t as a dummy variable to avoid a non-unique indefinite integral. In the second line, bringing the integral inside the sum is justified because we know from line 1 and Equation 3.3.25 the sum is uniformly convergent for

$x \in [0,1]$. The third line is obtained by performing the required integration in line 2 and noting the result is unique up to a constant C. The final line is an application of the logarithm property $\log \frac{a}{b} = \log a - \log b$. $\square$

3.3.2 Necessary Lemmas

In this section we will assume that a sequence of functions f_n satisfy a recurrent functional relationship similar to Equation 3.3.25, and by assuming this recurrence relationship, we will prove four main results governing the behavior of the f_n. These results will be crucial in showing f_n is bounded above and below by $a \pm Be^{-\lambda\sqrt{n}}$, where B and λ are positive constants and a will be defined later in this chapter. If we substitute $m_n^{'}(x)$ in for $f_n(x)$ and integrate both $m_n^{'}(x)$ and the bounds for $f_n(x)$, then Kuzmin's Theorem will follow.

For the following lemmas, we assume that we have an infinite sequence of real functions $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ defined on $x \in [0,1]$, satisfying the following conditions:

$$f_{n+1}(x) = \sum_{k=1}^{\infty} \frac{1}{(k+x)^2} f_n\Big(\frac{1}{k+x}\Big), \;\; 0 \leq n \tag{3.3.28}$$

where clearly $\frac{1}{k+x}$ is the argument of f_i; the sequence of f_i also satisfies:

$$0 < f_0(x) < M \;\text{ and }\; |f_0^{'}(x)| < \tau. \tag{3.3.29}$$

The series in Equation 3.3.28 is uniformly convergent due to our analysis of Equation 3.3.25.

Lemma 3.3.3. *For any* $0 \leq n$ *and* $x \in [0,1]$*, we have the following:*

$$f_n(x) = \sum^{(n)} f_0\Big(\frac{p_n + xp_{n-1}}{q_n + xq_{n-1}}\Big)\frac{1}{(q_n + xq_{n-1})^2}, \tag{3.3.30}$$

where (n) *denotes the sum over all intervals of rank* n*,* $\left(\frac{p_n}{q_n}, \frac{p_n+p_{n-1}}{q_n+q_{n-1}}\right)$ *is an arbitrary interval of rank* n*, and* $\frac{p_n+xp_{n-1}}{q_n+xq_{n-1}}$ *is the argument of* f_0*.*

Proof: This proof proceeds by induction. First we establish the base case: let $n = 0$, then $f_0(x) = \sum^{(0)} f_0\Big(\frac{p_0+xp_{-1}}{q_0+xq_{-1}}\Big)\frac{1}{(q_0+xq_{-1})^2}$; the sum is over all rank 0 intervals, which coincides the single interval $[0,1]$. Thus, we have $p_0 = 0, \; q_0 =$

$1,\ p_{-1} = 1,\ q_{-1} = 0$. Now, assume that this relationship holds for all $i \leq n$. Then we proceed from Equation 3.3.28:

$$\begin{aligned} f_{n+1}(x) &= \sum_{k=1}^{\infty} \frac{1}{(k+x)^2} f_n\Big(\frac{1}{k+x}\Big) \\ &= \sum_{k=1}^{\infty} \frac{1}{(k+x)^2} \sum^{(n)} f_0\left(\frac{p_n + \frac{1}{k+x}p_{n-1}}{q_n + \frac{1}{k+x}q_{n-1}}\right) \frac{1}{(q_n + \frac{1}{k+x}q_{n-1})^2} \\ &= \sum^{(n)} \sum_{k=1}^{\infty} f_0\Big(\frac{(p_n k + p_{n-1}) + x p_n}{(q_n k + q_{n-1}) + x q_n}\Big) \frac{1}{\{(q_n k + q_{n-1}) + x q_n\}^2} \\ &= \sum^{(n+1)} f_0\Big(\frac{p_{n+1} + x p_n}{q_{n+1} + x q_n}\Big) \frac{1}{(q_{n+1} + x q_n)^2}, \end{aligned} \tag{3.3.31}$$

where the second equation results from a substitution based on our inductive assumption that the relationship in Equation 3.3.30 holds for all $i \leq n$. The third line is a result of multiplying the term $\frac{1}{(q_n + \frac{1}{k+x}q_{n-1})^2}$ by $\frac{1}{(k+x)^2}$ in line 2, expanding all numerators and denominators, and then grouping the terms. We also switch the order of summation because Equation 3.3.28 is uniformly convergent for $0 \leq x \leq 1$; thus, extending the equality sign from the first line in Equation 3.3.31, we conclude that the series in line 3 is also uniformly convergent. Finally, taking the sum in line 3 over k from 1 to ∞, which results in an arbitrary interval of rank $n+1$, yields the fourth line. The sum in line 4 is over all intervals of rank $n+1$ because we previously summed over all $k \in N$, which yielded all possible intervals of rank $n+1$ within each interval of rank n. Since the union of all intervals of rank $n+1$ cover all the intervals of rank n, the sum in line 4 is justified.□

We will now present a lemma bounding $|f_n'(x)|$ beyond the initial condition in Equation 3.3.29.

Lemma 3.3.4. *Given Equations 3.3.28 and 3.3.29, we have*

$$|f_n'(x)| < \frac{\tau}{2^{n-3}} + 4M. \tag{3.3.32}$$

Proof: We have from the last lemma that

$$f_n(x) = \sum^{(n)} f_0\Big(\frac{p_n + x p_{n-1}}{q_n + x q_{n-1}}\Big) \frac{1}{(q_n + x q_{n-1})^2}, \tag{3.3.33}$$

and we know from our argument above that this series is uniformly convergent for $0 \leq x \leq 1$, so differentiating the series term by term resuls in:

$$f_n'(x) = \sum^{(n)} f_0'(u) \frac{(-1)^{n-1}}{(q_n + xq_{n-1})^4} - 2\sum^{(n)} f_0(u) \frac{q_{n-1}}{(q_n + xq_{n-1})^3} \qquad (3.3.34)$$

where we let $u = \frac{p_n + xp_{n-1}}{q_n + xq_{n-1}}$, and we note by Lemma 1.4.3 that ($p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$).

Following the analysis in both Deveaux [De] and Khintchine[Ki], we conclude that ($\frac{q_{n-1}}{(q_n + xq_{n-1})^3} \leq \frac{q_{n-1}}{q_n} \frac{1}{q_n^2} < \frac{1}{q_n^2}$) because $q_{n-1} < q_n$. Also, we have as a consequence of Theorem 1.4.2, ($q_n > q_{n-1} \Rightarrow q_n > \frac{q_n + q_{n-1}}{2} \Rightarrow q_n^2 > \frac{q_n(q_n + q_{n-1})}{2}$) (where we have multiplied each side by q_n). Therefore, by substituting the appropriate inequality, we can bound the second term on the right hand side of Equation 3.3.34:

$$\begin{aligned} \left| 2\sum^{(n)} f_0(u) \frac{q_{n-1}}{(q_n + xq_{n-1})^3} \right| &< \left| 2\sum^{(n)} f_0(u) \frac{1}{(q_n)^2} \right| \\ &< \left| 4\sum^{(n)} f_0(u) \frac{1}{q_n(q_n + q_{n-1})} \right| \\ &< \left| 4M\sum^{(n)} \frac{1}{q_n(q_n + q_{n-1})} \right| = 4M, \end{aligned} \qquad (3.3.35)$$

where last inequality results from our condition in Equation 3.3.29, namely ($0 < f_0(x) < M$). The last step is a consequence of $\sum^{(n)} \frac{1}{q_n(q_n + q_{n-1})} = \sum^{(n)} \left| \frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}} \right| = 1$. This sum equals 1 because we are summing over all intervals of rank n, which cover $[0, 1]$, see Equation 3.2.13.

Now considering the second term in Equation 3.3.34, we note from Theorem 1.4.2 that ($q_n \geq q_{n-1} + q_{n-2} \geq 2q_{n-2}$) and $q_1 = 1$; by repeated application of this inequality, we have ($q_n(q_n + q_{n-1}) > q_n^2 > 2^{n-1}$) (by induction the denominators of the convergents increase by at least a factor of 2 for each $n > 1$). As a result,

$$
\begin{aligned}
(q_n + xq_{n-1})^4 > q_n^4 \quad &> \quad \left(\frac{q_n(q_n+q_{n-1})}{2}\right)^2 \\
= \frac{q_n(q_n+q_{n-1})}{4}\Big(q_n(q_n+q_{n-1})\Big) \quad &> \quad 2^{n-3} q_n(q_n+q_{n-1}) \\
\Rightarrow \Big|\sum^{(n)} f_0'(u)\frac{-1}{(q_n+xq_{n-1})^4}\Big| \quad &< \quad \frac{\tau}{2^{n-3}}.
\end{aligned} \tag{3.3.36}
$$

The second inequality is a result of the relation $q_n > \frac{q_n+q_{n-1}}{2}$. The last inequality results from lines 1 and 2, Equation 3.2.13, and our condition in Equation 3.3.29, which gives $|f_0'(x)| < \tau$. The lemma follows. □

The next two lemmas are relatively straightforward and require only a few lines to prove.

Lemma 3.3.5. *For* $(0 \leq x \leq 1)$, *if*

$$
\frac{t}{1+x} < f_n(x) < \frac{T}{1+x} \tag{3.3.37}
$$

then we also have

$$
\frac{t}{1+x} < f_{n+1}(x) < \frac{T}{1+x} \tag{3.3.38}
$$

[Ki].

Proof: By Equation 3.3.28 and the assumption $\frac{t}{1+x} < f_n(x) < \frac{T}{1+x}$ from this lemma, we reason that:

$$
\begin{aligned}
\sum_{k=1}^{\infty} \frac{t}{1+\frac{1}{k+x}}\frac{1}{(k+x)^2} < \quad f_{n+1}(x) \quad &< \sum_{k=1}^{\infty} \frac{T}{1+\frac{1}{k+x}}\frac{1}{(k+x)^2} \\
= t\sum_{k=1}^{\infty} \frac{1}{(k+x)(k+x+1)} < \quad f_{n+1}(x) \quad &< T\sum_{k=1}^{\infty} \frac{1}{(k+x)(k+x+1)} \\
= t\sum_{k=1}^{\infty}\left(\frac{1}{k+x} - \frac{1}{k+x+1}\right) < \quad f_{n+1}(x) \quad &< T\sum_{k=1}^{\infty}\left(\frac{1}{k+x} - \frac{1}{k+x+1}\right) \\
\Rightarrow \frac{t}{1+x} < \quad f_{n+1}(x) \quad &< \frac{T}{1+x}.
\end{aligned} \tag{3.3.39}
$$

Line 1 is a direct result of our condition in Equation 3.3.28. The second line is obtained by multiplying and expanding the sums in line 1. Separating the terms being summed in line 2 into their partial fraction representations gives line 3. Evaluating the telescoping series in line 3 yields line 4. □

We now present the final lemma.

Lemma 3.3.6. *For all integer values* $n \geq 0$, *we have:*

$$\int_0^1 f_n(z)\,dz = \int_0^1 f_0(z)\,dz \quad [Ki]. \tag{3.3.40}$$

Proof: This is proof follows by induction. Let $n = 0$, then we obtain $\int_0^1 f_0(z)\,dz = \int_0^1 f_0(z)\,dz$. Now, assume the relation in Equation 3.3.40 holds for all $i < n$. Then we have:

$$\begin{aligned} \int_0^1 f_n(z)\,dz &= \sum_{k=1}^{\infty} \int_0^1 f_{n-1}\Big(\frac{1}{k+z}\Big)\frac{dz}{(k+z)^2} \\ = \sum_{k=1}^{\infty} \int_{\frac{1}{k+1}}^{\frac{1}{k}} f_{n-1}(u)\,du &= \int_0^1 f_{n-1}(u)\,du = \int_0^1 f_0(z)\,dz, \end{aligned} \tag{3.3.41}$$

for all $n > 0$. The first line is basically a restatement of Equation 3.3.28, where we switch the intergral and the sum because the sum is uniformly convergent. We let $u = \frac{1}{k+z}$ in order to proceed from line 1 to line 2. Summing $\int_{\frac{1}{k+1}}^{\frac{1}{k}} f_{n-1}(u)\,du$ over all k justifies the third equality (formally, we should write $\lim_{b \to 0} \int_b^1 f_{n-1}(u)\,du$). Our inductive assumption justifies our final equality. Thus our lemma is proved. □

3.3.3 Proof of Main Result

If a function f is strictly positive and continuous in a closed interval $[a, b]$, then f possesses a positive minimum $f(x) = m$ for some $x \in [a, b]$. We assumed in

Equation 3.3.29 that $f_0'(x)$ exists and $0 < f_0(x) < M$ for all $x \in [0,1]$. Therefore, since $0 < f_0(x)$ and differentiability implies continuity, we conclude that $f_0(x)$ possesses a positive minimum m for some $x \in [0,1]$. Thus, we argue that for $x \in [0,1]$:

$$\begin{aligned} m \leq \quad & f_0(x) \quad < M \\ \frac{m}{2(1+x)} < \quad & f_0(x) \quad < \frac{2M}{1+x} \\ \frac{g}{1+x} < \quad & f_0(x) \quad < \frac{G}{1+x}, \end{aligned} \tag{3.3.42}$$

where the first step summarizes our assumptions. To obtain line 2, we note $x > 0$, so we can divide line 1 by $(1+x)$ and introduce a factor of $\frac{1}{2}$ and 2 to ensure strict inequalities because $1 \leq (1+x) \leq 2$. In line 3 we let $g = \frac{m}{2}$ and $G = 2M$. It is important to note we hinge the following line of analysis on the base function $f_0(x)$ and not $f_n(x)$.

Now we define a function that is strictly positive, is defined for all non-negative integer values of n, and whose domain is $x \in [0,1]$:

$$\phi_n(x) = f_n(x) - \frac{g}{1+x} \tag{3.3.43}$$

Recall from Equation 3.3.27 that the function $\theta(x) = C\ln(1+x)$ satisfied the relationship $\sum\limits_{k=1}^{\infty}\left(\theta\left(\frac{1}{k}\right) - \theta\left(\frac{1}{k+x}\right)\right)$, which we know upon differentiation satisfies $\theta'(x) = \sum\limits_{k=1}^{\infty} \frac{1}{(k+x)^2}\theta'\left(\frac{1}{k+x}\right)$. Motivated by these relationships, we define $F(x) = \frac{g}{1+x}$, which is the derivative of $g\ln(1+x)$; therefore:

$$F(x) = \sum_{k=1}^{\infty} F\left(\frac{1}{k+x}\right)\frac{1}{(k+x)^2}. \tag{3.3.44}$$

By the definition provided in Equation 3.3.28, f_n satisfies the same functional relationship as $F(x)$ does in Equation 3.3.44, and as a result, the sequence of functions $\phi_0(x), \phi_1(x), \ldots, \phi_n(x), \ldots$ satisfies this same functional relationship. Khintchine makes the astute observation that because the sequence of functions $\phi_0(x), \phi_1(x), \ldots, \phi_n(x), \ldots$ satisfies the relationship in Equation 3.3.28, all the lemmas presented in the subsection "Necessary Lemmas" hold for this sequence, and in particular, each function $\phi_i(x)$ satisifies the relationship established in Equation 3.3.30.

Our next goal is to bound $\phi_n(x)$ from above and below by functions of n. If we can find such bounding functions, then we can show the bounds for the sequence of $f_n(x)$ converge to the same value.

Recalling that $u = \frac{p_n+xp_{n-1}}{q_n+xq_{n-1}}$, we can rewrite $\phi_n(x)$ with the aid of Lemma 3.3.3:

$$\phi_n(x) = \sum^{(n)} \phi_0(u) \frac{1}{(q_n + xq_{n-1})^2}, \tag{3.3.45}$$

and since $x \leq 1$ and $q_n \geq q_{n-1} + q_{n-2}$ we have

$$q_n + xq_{n-1} \leq q_n + q_{n-1} < 2q_n. \tag{3.3.46}$$

From Equation 3.3.42 and from the definition of $\phi_0(u)$ given in Equation 3.3.43, it is clear $\phi_0(u) > 0$ because $f_0(x) > \frac{g}{1+x}$; we can apply Lemma 3.3.5 to conclude $\phi_n > 0$ for all n, therefore

$$\frac{1}{2} \sum^{(n)} \phi_0(u) \frac{1}{q_n(q_n + q_{n-1})} < \phi_n(x) \tag{3.3.47}$$

Substituting Equation 3.3.45 in 3.3.46, namely $(q_n+xq_{n-1})^2 = (q_n+xq_{n-1})(q_n+xq_{n-1}) < 2q_n(q_n + q_{n-1})$, gives Equation 3.3.47.

The Mean Value Theorem from real analysis allows us to write:

$$\begin{aligned} \int_{\frac{p_n}{q_n}}^{\frac{p_n+p_{n-1}}{q_n+q_{n-1}}} \phi_0(z)\,dz &= \phi_0(u_n^{'}) \frac{1}{q_n(q_n + q_{n-1})} \\ \Rightarrow \frac{1}{2} \int_0^1 \phi_0(z)\,dz &= \frac{1}{2} \sum^{(n)} \phi_0(u_n^{'}) \frac{1}{q_n(q_n + q_{n-1})}. \end{aligned} \tag{3.3.48}$$

Here, we note $u_n^{'} \in \left(\frac{p_n}{q_n}, \frac{p_n+p_{n-1}}{q_n+q_{n-1}}\right)$, and we apply the Mean Value Theorem to every disjoint rank n interval (i.e. $I_{(a_1,\ldots,a_n)} = [\frac{p_n}{q_n}, \frac{p_n+p_{n-1}}{q_n+q_{n-1}}]$). Because we apply the Mean Value Theorem to every rank n interval, $u_n^{'}$ is different for each rank n

interval, over which we integrate ϕ_0. The second equation is a result of summing both sides of the equality in line 1 over all possible intervals of rank n. The right hand side of line 2 is an expression of the Riemann sum, where we have $\phi_0(u_n^{'})$ as the representative height in a particular interval of rank n and $\frac{1}{q_n(q_n+q_{n-1})}$ is the length of each rank n interval.

Combining Equations 3.3.47 and 3.3.48, we arrive at the inequality:

$$\phi_n(x) - \frac{1}{2}\int_0^1 \phi_0(z)\,dz > \frac{1}{2}\sum^{(n)}\{\phi_0(u) - \phi_0(u^{'})\}\frac{1}{q_n(q_n+q_{n-1})}. \tag{3.3.49}$$

Then if we differentiate $\phi(x)$ in Equation 3.3.43, while keeping in mind that by Equation 3.3.29 we have $|f_0^{'}(x)| < \tau$ and that $\left|(\frac{g}{1+x})^{'}\right| = \left| - \frac{g}{(1+x)^2}\right| \leq g$, we conclude for $x \in [0,1]$:

$$|\phi_0^{'}(x)| \leq |f_0^{'}(x)| + g < \tau + g \tag{3.3.50}$$

In Equation 3.3.48 we established $u_n^{'} \in \left(\frac{p_n}{q_n}, \frac{p_n+p_{n-1}}{q_n+q_{n-1}}\right)$, therefore:

$$|u - u^{'}| < \frac{1}{q_n(q_n+q_{n-1})} < \frac{1}{q_n^2} < \frac{1}{2^{n-1}}. \tag{3.3.51}$$

Combining this equation with Equation 3.3.50, we infer:

$$\begin{aligned}
\frac{|\phi_0(u_n) - \phi_0(u_n^{'})|}{|u_n - u_n^{'}|} = \phi_0^{'}(u_n^{'}) &< \tau + g \\
\Rightarrow |\phi_0(u_n) - \phi_0(u_n^{'})| &< (\tau + g)|u_n - u_n^{'}| \\
< \frac{\tau+g}{q_n(q_n+q_{n-1})} &< \frac{\tau+g}{q_n^2} \\
&< \frac{\tau+g}{2^{n-1}}.
\end{aligned} \tag{3.3.52}$$

This argument is straightforward except for proceeding from line 2 to line 3; since $u_n^{'} \in \left(\frac{p_n}{q_n}, \frac{p_n+p_{n-1}}{q_n+q_{n-1}}\right)$, which denotes a rank n interval having length $\frac{1}{q_n(q_n+q_{n-1})}$, the inequality $(\tau+g)|u_n - u_n^{'}| < \frac{\tau+g}{q_n(q_n+q_{n-1})}$ holds for all $u_n^{'}$. Thus, combining Equations 3.3.49 and 3.3.52, we have:

$$\phi_n(z) > \frac{1}{2}\int_0^1 \phi_0(z)\,dz - \frac{\tau+g}{2^n} = l - \frac{\tau+g}{2^n} \tag{3.3.53}$$

where,

$$l = \frac{1}{2}\int_0^1 \phi_0(z)\,dz \tag{3.3.54}$$

Collecting our results and recalling Equation 3.3.43, we conclude:

$$f_n(x) > \frac{g}{1+x} + l - \frac{\tau+g}{2^n} > \frac{g + l - 2^{-n+1}(\tau+g)}{1+x} = \frac{g_1}{1+x} \tag{3.3.55}$$

Now, consider a new sequence of functions defined for all non-negative integers n and for all $x \in [0,1]$:

$$\sigma_n(x) = \frac{G}{1+x} - f_n(x), \tag{3.3.56}$$

Applying the same logic used to obtain Equation3.3.55, we obtain an upper bound for the sequence of functions $\sigma_0,\ \sigma_1,\ \sigma_2, \ldots$ defined in Equation 3.3.56:

$$f_n(x) < \frac{G - l^{'} + 2^{-n+1}(\tau+G)}{1+x} = \frac{G_1}{1+x} \tag{3.3.57}$$

where $l^{'} = \frac{1}{2}\int_0^1 \sigma_0(z)\,dz$. We have thus established the upper and lower bounds of f_n as functions of n (i.e. G_1, g_1 are functions of n), which is what we initially set out to find.

Subtracting Equation 3.3.55 from Equation 3.3.57 and realizing l, $l^{'} > 0$, we have $g < g_1 < G_1 < G$. These inequalities apply for large n, or as $\lim_{n\to\infty} 2^{-n+1} = 0$. As a result:

$$G_1 - g_1 < G - g - (l + l^{'}) + 2^{-n+2}(\tau+G) \tag{3.3.58}$$

Then using the definitions of l and $l^{'}$ and the definitions of ϕ and σ introduced in Equations 3.3.43 and 3.3.56:

$$l + l^{'} = \frac{1}{2}\int_0^1 \frac{G-g}{1+z}\,dz = (G-g)\frac{\ln(2)}{2}, \tag{3.3.59}$$

where G and g are constants. Combining our previous result with Equation 3.3.58, we infer:

$$G_1 - g_1 < (G - g)\delta + 2^{-n+2}(\tau + G), \tag{3.3.60}$$

where $\delta = 1 - \frac{\ln(2)}{2} < 1$, which is a positive constant. In summary, we note from Equation 3.3.42:

$$\frac{g}{1+x} < f_0(x) < \frac{G}{1+x}, \tag{3.3.61}$$

and for large enough n, we just concluded:

$$\frac{g_1}{1+x} < f_n(x) < \frac{G_1}{1+x}, \tag{3.3.62}$$

where all the relations among G, g, G_1, g_1 in Equation 3.3.58 still hold.

In the beginning of this line of analysis, we commented that our base function was f_0. Suppose we considered $f_n(x)$ as our starting function and reapplied the same rigorous argument to this function, then it is evident:

$$\frac{g_2}{1+x} < f_{2n}(x) < \frac{G_2}{1+x}, \tag{3.3.63}$$

where we have $G_2 - g_2 < (G_1 - g_1)\delta + 2^{-n+2}(\tau_1 + G_1)$ and $g_1 < g_2 < G_2 < G_1$ (see Equation 3.3.60), and τ_1 is defined in a similar fashion to Equation 3.3.29, namely $|f_n'(x)| < \tau_1$. Now, we can continue this process an infinite number of times, which produces a general result:

$$\frac{g_r}{1+x} < f_{rn}(x) < \frac{G_r}{1+x}, \tag{3.3.64}$$

which naturally implies a relationship among G_is and g_is similar to Equation 3.3.60:

$$G_r - g_r < (G_{r-1} - g_{r-1})\delta + 2^{-n+2}(\tau_{r-1} + G_{r-1}), \tag{3.3.65}$$

where this equation yields the relationship $g_{r-1} < g_r < G_r < G_{r-1}$. Again, τ_{r-1} is defined in a similar fashion to Equation 3.3.29, namely $|f_{(r-1)n}'(x)| < \tau_{r-1}$. **The above inequalities hold for** $r \in N$ **and** $x \in [0, 1]$.

By Lemma 3.3.4, we can write $\tau_r < \frac{\mu}{2^{rn-3}} + 4M$ for $r \in N$, from which we see $\lim\limits_{n\to\infty} \frac{\tau}{2^{rn-3}} = \infty$; thus $\tau_r < 5M$ for large n and $r \in N$. We can repeat the application of Equation 3.3.65 for all $r \in [1, n]$:

$$G_n - g_n < (G-g)\delta^n + 2^{-n+2}\{(\mu+2M)\delta^{n-1} + 7M\delta^{n-1} + 7M\delta^{n-3} + \ldots + 7M\delta + 7M\}, \tag{3.3.66}$$

which expresses $G_n - g_n$ in terms of G and g. Because both $\delta < 1$ and 2^{-n+2} decay very rapidly as $n \to \infty$, we can bound this decay from above with e^{-n}, in particular:

$$G_n - g_n < Be^{-\lambda n}. \tag{3.3.67}$$

Since $G_n - g_n$ is a function of τ and M, and G, g are functions of M, B must be a function of both M and τ, formally $B = B(M, \tau)$. Since the expression in Equation 3.3.66 is strictly positive, $B > 0$ for all M and τ. Finally, λ is an absolute constant that does not change irrespective of the sequence of functions f_n satisfying Equation 3.3.28, and $\lambda < 1$ otherwise for any $B \in R$ as $n \to \infty$, we would have $Be^{-\lambda n} < 2^{-n+2}$, thereby violating the assumption needed for Equation3.3.67.

From Equation 3.3.67, we complete our goal of showing the lower and upper bounds of f_n converge to the same value as $n \to \infty$:

$$\lim_{n\to\infty} G_n = \lim_{n\to\infty} g_n = a, \tag{3.3.68}$$

where this limit exists because $\lim\limits_{n\to\infty} Be^{-\lambda n} = 0$. Now, let $r = n$ in Equation 3.3.64, and we argue for $x \in [0, 1]$:

$$\left| f_{n^2}(x) - \frac{a}{1+x} \right| < Be^{-\lambda n}. \tag{3.3.69}$$

If we considered this inequality as $n \to \infty$, then it is clear the sequence of f_n converges uniformly to $\frac{a}{1+x}$. Using Equation 3.3.69 and a result in real analysis that states if a sequence is uniformly convergent, then the limit of the integrals is the integral of the limit, we reason:

$$\begin{aligned} \lim_{n\to\infty} f_{n^2}(x) &= \frac{a}{1+x} \\ \Rightarrow \lim_{n\to\infty} \int_0^1 f_{n^2}(z)\,dz &\to a\ln 2. \end{aligned} \tag{3.3.70}$$

Hence, Lemma 3.3.6 gives $a = \frac{1}{\ln 2} \int_0^1 f_0(z)\, dz$. If we now let N be such that $n^2 \leq N < (n+1)^2$, where n^2 is the index used in the sequence f_{n^2}, then dropping the absolute value signs in Equation 3.3.69 yields:

$$\begin{aligned} \frac{a - 2Be^{-\lambda n}}{1+x} < f_{n^2}(x) &< \frac{a + 2Be^{-\lambda n}}{1+x} \\ \Rightarrow \frac{a - 2Be^{-\lambda n}}{1+x} < f_N(x) &< \frac{a + 2Be^{-\lambda n}}{1+x} \\ \Rightarrow \left| f_N(x) - \frac{a}{1+x} \right| < 2Be^{-\lambda n} &= Ae^{-\lambda(n+1)} < Ae^{-\lambda\sqrt{N}}, \end{aligned} \quad (3.3.71)$$

where $A = 2Be^{\lambda}$. Note, line 2 is the result of applying Lemma 3.3.5 to f_{n^2} in line 1. The first inequality in line 3 is the consequence of introducing absolute value into line 2, and the final inequality is due to our construction of $N < (n+1)^2$.

Throughout this section, our results have held for sufficiently large N, which implies Equation 3.3.71 holds only for large N, but if A can be made arbitrarily large, then Equation 3.3.71 can be made to hold for all $N \geq 0$. The arguments presented in this section have proved the following theorem.

Theorem 3.3.7. ***Kuzmin's Theorem*** *- Let the conditions of Equations 3.3.28 and 3.3.29 hold, then*

$$f_n(x) = \frac{a}{1+x} + \theta A e^{-\lambda\sqrt{n}}, \quad (3.3.72)$$

where $x \in [0,1]$, $a = \frac{1}{\ln 2} \int_0^1 f_0(z)\, dz$, $|\theta| < 1$, $\lambda < 1$ *is an absolute positive constant, and* A *is a positive function of* M *and* τ *but not of* x *[Ki].*

A more detailed analysis of the constants in Kuzmin's Theorem will be provided at the end of this chapter; however, we will find A and λ cannot be assigned actual numerical values without losing some of the theoretical thrust of the theorem. Kuzmin's Theorem implies a result that is paramount to our analysis that is presented at the end of this chapter.

3.3.4 Kuzmin's Result

Recall the definition of $m_n(x)$ and Equations 3.3.21 and 3.3.25, which show:

$$\begin{aligned} m_{n+1}(x) &= \sum_{k=1}^{\infty}\left(m_n\Big(\frac{1}{k}\Big) - m_n\Big(\frac{1}{k+x}\Big)\right) \\ m'_{n+1}(x) &= \sum_{k=1}^{\infty}\frac{1}{(k+x)^2}m'_n\Big(\frac{1}{k+x}\Big). \end{aligned} \tag{3.3.73}$$

But in Equation 3.3.28, we defined $f_{n+1}(x)$ to satisfy:

$$f_{n+1}(x) = \sum_{k=1}^{\infty}\frac{1}{(k+x)^2}f_n\Big(\frac{1}{k+x}\Big), \tag{3.3.74}$$

which is precisely $m'_{n+1}(x)$'s functional relationship. Also note ($\int m'_{n+1}(x)\,dx = m_{n+1}(x)$), which is the measure of the set of numbers α in the interval $[0,1]$ for which $z_n(\alpha) < x$.

Gauss wanted to find a closed form expression for $m_n(x)$ for large n. Motivated by the similarities between Equations 3.3.73 and 3.3.74, we set:

$$f_n(x) = m'_n(x) \tag{3.3.75}$$

for $x \in [0,1]$. If we let $f_0(x) \equiv 1$, then all the conditions of Theorem 3.3.7 are satisfied. Now, we can appropriately apply Theorem 3.3.7 with $f_n(x) = m'_n(x)$:

$$\left|m'_n(x) - \frac{1}{(1+x)\ln 2}\right| < Ae^{-\lambda\sqrt{n}} \tag{3.3.76}$$

integrating yields

$$\left|m_n(x) - \frac{\ln(1+x)}{\ln 2}\right| < Ae^{-\lambda\sqrt{n}}, \tag{3.3.77}$$

where these inequalities hold for $x \in [0,1]$. Again, A and λ are absolute positive constants. Note, the factor of x, which should result from integrating the right hand side with respect to x, is dropped. We drop the x in order to make the error term independent of x, and we note $xAe^{-\lambda\sqrt{n}} \le Ae^{-\lambda\sqrt{n}}$ because $x \in [0,1]$. Gauss's conjecture is thus proved.

We can apply these results to approximate the measure of the set of numbers for which $a_n = k$ for large n.

Recall from Equation 3.3.22 $z_n = r_n - a_n = [0; a_{n+1}, \ldots]$, and as a result, if $a_n = k$, then

$$\frac{1}{k+1} < z_{n-1}(\alpha) \leq \frac{1}{k}; \tag{3.3.78}$$

therefore

$$\mu(E\binom{n}{k}) = m_{n-1}\left(\frac{1}{k}\right) - m_{n-1}\left(\frac{1}{k+1}\right) = \int\limits_{\frac{1}{k+1}}^{\frac{1}{k}} m_{n-1}^{'}(x)\,dx. \tag{3.3.79}$$

Now, we integrate Equation 3.3.77 from $\frac{1}{k+1}$ to $\frac{1}{k}$ and use Equation 3.3.79:

$$\left|\mu(E\binom{n}{k}) - \frac{\ln\left\{1+\frac{1}{k(k+2)}\right\}}{\ln 2}\right| < \frac{A}{k(k+1)}e^{-\lambda\sqrt{n-1}}, \tag{3.3.80}$$

where the power $\sqrt{n-1}$ is a result Equation 3.3.78. The factor $C\frac{1}{k(k+1)}$ is a result of the integration $= \int\limits_{\frac{1}{k+1}}^{\frac{1}{k}} C\,dx$, where $C = Ae^{-\lambda\sqrt{n-1}}$ is not a function of x. Note, $\frac{1}{k(k+1)}$ is the length of each interval characterized by $a_n = k$. Collecting our results, we conclude this section with Kuzmin's result:

$$\mu(E\binom{n}{k}) \rightarrow \frac{\ln\left\{1+\frac{1}{k(k+2)}\right\}}{\ln 2} \tag{3.3.81}$$

as $n \rightarrow \infty$.

3.4 Levy's Refined Results

DeVeaux [De] provides an excellent summary of Levy's proof, which DeVeaux uses to critique Kuzmin's approach. The major criticism of Kuzmin's proof is the reliance on the seemingly too restrictive condition in Equations 3.3.28, "which shows convergence regardless of the distribution chosen for X [De]." Levy did not rely on such heavy assumptions, which enabled him to solve Gauss' problem not only for x uniformly distributed in $[0, 1]$, but also for x with a density in the set of Lebesgue measurable functions ($L^1[0, 1]$).

Khintchine states: "The method of P. Levy allows a better estimate to be obtained. The inequality

$$\left| m_n(x) - \frac{\ln(1+x)}{\ln 2} \right| < Ae^{-\lambda n} \tag{3.4.82}$$

is shown to be satisfied [Ki]." If we substitute this error term into the inequality in Equation 3.3.77 and follow the exact same argument used to obtain Equation 3.3.80, then we conclude:

$$\Rightarrow \left| \mu(E \binom{n}{k}) - \frac{\ln\left\{1 + \frac{1}{k(k+2)}\right\}}{\ln 2} \right| < \frac{A}{k(k+1)} e^{-\lambda(n-1)}. \tag{3.4.83}$$

Because $e^{-\lambda(n-1)} \to 0$ much faster than $e^{-\lambda\sqrt{n-1}} \to 0$ as $n \to \infty$, we conclude Levy's bounds are significantly better than Kuzmin's, but still not necessarily optimal.

3.5 Experimental Results for Levy's Constants

3.5.1 Motivation

Does there exists a more optimal bounding function for the difference $\left| \mu(E \binom{n}{k}) - \frac{\ln\left\{1+\frac{1}{k(k+2)}\right\}}{\ln 2} \right|$ than Levy's $\frac{A}{k(k+1)} e^{-\lambda n}$? Formally, can we find a function $e^{g(n)}$, such that $e^{g(n)} < e^{-\lambda n}$ and $\left| \mu(E \binom{n}{k}) - \frac{\ln\left\{1+\frac{1}{k(k+2)}\right\}}{\ln 2} \right| < CA(k)e^{g(n)} = \theta(n,k)$ holds for all n or at least for large n? Intuitively, one may believe there exist a more optimal bounding function $\theta(n,k)$ for all n than Levy's proposed function, even if the order of $\theta(n,k)$ is equal only to e^{-cn}; or, $g(n)$ has a higher order than cn, which is the order of Levy's bounding function (i.e. $e^{-\lambda(n-1)}$).

This intuition results from two facts. First, Levy's function does not bound the difference in Equation 3.4.83 with equality, so perhaps there exists a bounding function that does indeed bound this difference with equality. Secondly, in going from line 1 to line 2 in Equations 3.3.77 and 3.4.83, we drop the factor of $x \leq 1$ from the right hand side. If this factor were included in Levy's error term, then not only would the error term depend on x, but also as $x \leq 1$ the resultant error term would be more optimal that our current error term in Equation 3.4.83.

Suppose we set out to emprirically estimate the optimal error function $\theta(n, k)$ by considering the various length expansions of a set of irrational $\alpha \in [0, 1]$. Immediately, we note this test set has zero measure, so we could never be certain $\theta(n, k)$ is the optimal error function for a set of $\alpha \in [0, 1]$ with positive measaure. However, if we just consider the various length expansions of our test α, then we can find a function $\theta(n, k)$ that approximates the difference in Equation 3.4.83 to a better degree than e^{-n}; $\theta(n, k)$ would preserve the inequality in Equation 3.4.83 only for those values of n and α in out test set. Furthermore, if one uses Kuzmin's Theorem in an empirical analysis, then the bounding function does not need to hold necessarily for all n, rather only over a range of n; however, over this range of n, $\theta(n, k)$ needs to preserve the inequality in Equation 3.4.83 for all examined k and α.

Alternatively, we can assume $\theta(n, k)$ has the same functional form as Levy's bounding function:

$$\theta(n, k) = \frac{C}{k(k+1)} e^{-\lambda'(n-1)} \tag{3.5.84}$$

By assumption, $\theta(n, k)$ now looks almost exactly like Levy's bounding function. While the constants A and λ in Levy's bounding function are supposed to be absolute constants that preserve Inequality 3.4.83 for almost all α and for all n, we can change these constants to suit our purposes. Rarely do we need our bounding function to hold for all n, and rarely do we need this bounding function to hold for all $\alpha \in K$ (where K denotes the set for which Kuzmin's/Levy's Theorem holds).

It is extremely important to bear in mind that the "range of n" is determined by the demands of the empirical analysis; hereinafter, the previous statement will be assumed when refering to the range of n. We also define the "beginning n," which is the first n in the range, over which the bounding function must satisfy Inequality 3.5.86.

A very nice consequence of Equation 3.4.83 is:

$$Prob(a_n = k) \leq \log_2 \left(1 + \frac{1}{k(k+2)}\right) + \epsilon, \tag{3.5.85}$$

where ϵ is either Kuzmin's or Levy's error term. In order to approximate the constants A and λ in Equation 3.4.83, we must empirically estimate $\mu(\alpha \in [0, 1] : a_n = k)$, or $Prob(a_n = k)$. To perform this estimation we do the following: compute n coefficients of the continued fraction expansion of a given α_0 and count the number of coefficients a_i whose value is k; then we divide this number by the total number of coefficients examined (i.e. n): call this procedure "Kuzmin Test Procedure." It is important to note Kuzmin's Theorem gives us the probability $a_n = k$ for given n and k; however, we can test the expression in Kuzmin's Theorem by implementing the Kuzmin Test Procedure. Basically, the difference between the Kuzmin Test Procedure and Kuzmin's Theorem is the former tests the expected number of digits equal to k given n coefficients while the later gives $Prob(a_n = k)$. Both [MT] or [Mi] show the expected number of coefficients equal to k given n coefficients is $n\log_2(1+\frac{1}{k(k+2)}) + \epsilon(n,k)$, where $\epsilon(n,k)$ is the error term that decreases as $n \to \infty$. Therefore, Kuzmin's Theorem does lead directly to an expectation of the number of digits equal to k given n coefficients.

In the next few sections I will present a method for finding the optimal constants (i.e. A and λ) of the bounding function, where "optimal" is in reference to an arbitrary empirical analysis, and I will apply this method to a numerical application.

3.5.2 Problems in Estimating A and λ

Since Levy provides a more optimal bounding function than Kuzmin, we assume $\theta(n,k)$ has the same functional form as Levy's bounding function (see the condition presented in Equation 3.5.84). Then we have from Equation 3.4.83:

$$\left|\mu(E\binom{n}{k}) - \frac{\ln\left\{1+\frac{1}{k(k+2)}\right\}}{\ln 2}\right| < \theta(n,k) = \frac{A}{k(k+1)}e^{\lambda(n-1)}, \qquad (3.5.86)$$

where we relabeled the constants from Equation 3.5.84 (i.e. $CA(k) = \frac{A}{k(k+1)}$ and $\lambda^{'} = \lambda$).

We note an immediate problem in trying to estimate empirically the constants A and λ. Although by formal construction our constants A and λ are supposed to be independent of α, in any empirical analysis both constants will be functions of the tested α's. To see this dependence, consider an empirical test of Kuzmin's Theorem, where we perform Kuzmin Test Procedure for a predetermined set of coefficient values k and a test set of α. Since $a_i(\alpha)$ is a function of α and the expression

$$\left|\mu(\alpha : a_i = k) - \frac{\ln\left\{1+\frac{1}{k(k+2)}\right\}}{\ln 2}\right| \qquad (3.5.87)$$

is estimated by counting the number of coefficients a_i with value k, we observe A and λ are functions of α. Furthermore, both constants will also be functions of the range of n and the beginning n.

For example, suppose we chose $\alpha_0 \in K$, such that $a_1(\alpha_0) = 10^8$ and the subsequent digits are free to assume any values. Further suppose we are attempting to determine A and λ in Equation 3.5.86 by utilizing the Kuzmin Test Procedure over the first two digits of α_0. If we computed the constants A and λ based only on the coefficients a_1, a_2 (i.e. the range of $n = 2$ and the beginning $n = 1$), then A and λ will be extremely large because immediately there is a large divergence from Kuzmin's Theorem. However, we assumed that $\alpha_0 \in K$, which means that as $n \to \infty$ the divergence from Kuzmin's Theoreom $\to 0$. The point of tracing the effects of this $a_1 = 10^8$ problem (will also be referred to as the large digit problem) is to show that any empirical analysis needs to consider many α in its test set and to be conducted over a large number n of coefficients. Note: For conciseness, when we refer to Kuzmin's Theorem, we mean Kuzmin's Theorem with Levy's bounding function

To understand this problem we note that Kuzmin's expected value for $\mu(\alpha : a_n = 10^8) \leq \log_2(1 + \frac{1}{10^8 \times (10^8+2)}) + \epsilon$, which for all practical purposes is 0. But since our empirical estimation of $\mu(\alpha : a_n = 10^8)$ was based on only one test α_0 and on only two coefficients, we are led to believe by Kuzmin Test Procedure that $\mu(\alpha : a_n = 10^8) = \frac{1}{2}$. Thus, we conclude the difference in Equation 3.4.83 is approximated by:

$$\left| \mu(E \binom{n}{k}) - \frac{\ln \left\{ 1 + \frac{1}{10^8(10^8+2)} \right\}}{\ln 2} \right| = \frac{1}{2} - \epsilon \tag{3.5.88}$$

This divergence $\frac{1}{2} - \epsilon$ would yield a very large value of A, although I acknowledge the empirically estimated value of A would be somewhat attenuated by ultimately dividing A by $k(k+1)$ as in Equation 3.4.83. From Equations 3.3.80 and 3.4.83, this A theoretically should remain constant for all tested values of k, but we will see empirical results indicate otherwise. Additionally, this divergence value implies a larger value of λ than would be the case if our test considered a longer range of n.

Reconsider our $a_1 = 10^8$ problem over a longer range of n instead of only over a range of 2 coefficients as above; the effect of the coefficient $a_1 = 10^8$ on the divergence value will be diluted, and our contants A and λ will become smaller (assuming that there are not frequent occurrences of the coefficient value

10^8, which is a reasonable assumption given that we assumed $\alpha_0 \in K$). For example, suppose that we recompute $\mu(\alpha : a_i = 10^8)$ but let $n = 100,000$. We know $\alpha_0 \in K$, so we can assume that $a_i \neq 10^8$ for all $2 \leq i \leq 100,000$. Now the divergence is given by:

$$\left| \mu(E \binom{n}{k}) - \frac{\ln \left\{ 1 + \frac{1}{10^8(10^8+2)} \right\}}{\ln 2} \right| = \frac{1}{100,000} - \epsilon, \tag{3.5.89}$$

which would yield more optimal values for A and λ than the values determined when $n = 2$.

Therefore, if we compute the constants A and λ from an examination of only the first n (where n is assumed to be small) coefficients of only one test $\alpha_0 \in K$, then the bounding function $Ae^{-\lambda(n-1)}$ will be too great (i.e.

$$\left| \mu(\alpha : a_n = k) - \frac{\ln \left\{ 1 + \frac{1}{k(k+2)} \right\}}{\ln 2} \right| \ll Ae^{-\lambda(n-1)}) \tag{3.5.90}$$

for most ranges of n and most values of k to be of any practical use. Khintchine even notes if we choose A and λ sufficiently large, then we can make our bounding function hold for almost all $\alpha \in K$, for *all* n, but these large valued constants are of no practical use because they are not close to being optimal even for a small range of n, or for a small beginning n.

The natural follow up question to the previous discussion is what should be our minimum range of n and our lowest beginning value of n for an empirical analysis to approximate the optimal values of A and λ? The answer to this question depends on the problem that one wishes to solve. This paper will present a method for estimating the values of A and λ, such that

$$\left| \mu(E \binom{n}{k}) - \frac{\ln \left\{ 1 + \frac{1}{k(k+2)} \right\}}{\ln 2} \right| < \frac{A}{k(k+1)} e^{-\lambda(n-1)} \tag{3.5.91}$$

holds for a given range of n and for all test α, or more specifically for most values k over the given range of n. The range of n is determined by experimental demands, but my method can be applied to all required ranges of n and all test values of α and k.

3.5.3 Method

Again, given a test set of α and n coefficients, I will present the method for estimating the constants A and λ appearing in Levy's error term, which is a better approximation of the difference in Inequality 3.4.83 than Kuzmin's error term. I will often call the value of the difference in Inequality 3.4.83 the divergence from Kuzmin's Theorem.

In [Mi], the author computed the continued fraction expansions to the $n = 500,000^{th}$ coefficient of the cube roots of the first 100 primes and the first 100 primes greater than 10^7. For each tested α and k, the divergence from Kuzmin's expectation was computed. This empirical study suggests these cube roots are elements of the set K. The reason the author considered the two different sets of test α was to ensure one set of numbers was independent of the other. I refer the reader to [MT] for a more indepth analysis of this problem, but here I will offer an excerpt:

" If one studies say $x^3 - p = 0$, as we vary p the first few digits will often be the same. For example, the continued fractions for 100000007, 100000037 and 100000039 all begin $[179, 3, 1, 2, 5, 2]$. Consider a large number n_0. Primes near it can be written as $n_0 + x$ for x small. Then

$$\begin{aligned}(n_0 + x)^{\frac{1}{3}} &= n_0^{\frac{1}{3}} \cdot \left(1 + \frac{x}{n_0}\right)^{\frac{1}{3}} \\ &\approx n_0^{\frac{1}{3}} \cdot \left(1 + \frac{1}{3}\frac{x}{n_0}\right) \\ &= n_0^{\frac{1}{3}} + \frac{x}{3n_0^{\frac{2}{3}}}. \end{aligned} \tag{3.5.92}$$

If n_0 is a perfect cube, then for small x relative to n_0, these numbers will all have the same first few digits (and the first digit should be somewhat large). Thus, if we want to average over different roots, the first few digits are not independent; in many of the experiments, digits 50,000 to 1,000,000 were investigated: for roots of numbers of size 10^{10}, this was sufficient to see independent behavior (though ideally one should look at autocorrelations to verify this claim. Also, Kuzmin's theorem describes the behavior for n large; thus, it is worthwhile to throw away the first few digits so we only study regions where the error term is small." [MT]

Motivated by the results in [Mi], I considered five αs from the first 100 primes and five αs from the second set of primes with the greatest divergence from Kuzmin's Theorem over all tested values of k. The rationale for choosing only five from each set is that I am mainly trying to illustrate a method. Additionally, one could conjecture a number $\alpha \in K$ that exhibited the greatest divergence from Kuzmin's Theorem over a given range of n coefficients for a certain k would yield constants A and λ that bound from above the divergence of all other tested values of k for all the other tested αs over the same range. Note, α's large divergence from Kuzmin's Theorem given a range of n coefficients for a certain k, does not imply that α maintains a relatively large divergence as $n \to \infty$ for the same k, or even for other values of k. In a sense, the continued fractions of the low-divergence αs exhibited a faster convergence rate to αs' true value than the α with the greatest divergence. According to this intuition, we are testing essentially 200 numbers. The more α considered in an empirical analysis, the less dependent A and λ are on any particular α (except for the α with the greatest divergence from Kuzmin's Theorem).

I examined the cube roots of $\alpha = 79\ ,\ 167\ ,\ 223\ ,\ 251\ ,\ 307\ ,\ 10,000,357\ ,$ $10,001,221\ ,\ 10,001,237\ ,\ 10,001,567\ ,\ 10,001,643$ for $n = 7,070\ ,\ 50,000\ ,$ $100,000\ ,\ 150,000\ ,\ \ldots\ ,\ 2,000,000$. I then computed the number of coefficients that have values $k = 1$, 2, 3, 4, 5, 96, 97, 98, 99, 100 for each α for each value of n.

The rationale for choosing n in $50,000$ increments was to record the divergence at multiple values of n, but the lengths of the increments were arbitrary. The beginning $n = 7,070$ because according to Kuzmin's Theorem, the $Prob(a_i = 100) = \log_2(1 + \frac{1}{100\times(100+2)})$, which corresponds to observing one in every $7,070$ coefficients whose value is 100. Since we are fairly confident all the tested $\alpha \in K$ [Mi], for smaller values of n we expect to see a very low number of coefficients with values $k = 100$; thus we face the problem presented in the discussion where we assumed that $a_1 = 10^8$. However, letting $n = 7,070$ will dilute most of the effect of observing multiple occurrences of $a_i = 100$ for $i \leq 7,070$.

For example, suppose that for $\alpha_0 \in K$ we observed 3 coefficients such that $k_i = 100$, then we empirically estimate $\mu(\alpha : a_n = 100) = \frac{3}{7,070} \approx .0004$ versus Kuzmin's expectation of $\mu(\alpha : a_n = 100) = \frac{1}{7,070} = .0001$, yielding a divergence of $.0003$, which is not so large as to limit severely the optimality of resultant values of A and λ.

While all of the discussion about the $a_1 = 10^8$ problem has been to provide warning that the resultant bounding function will not be optimal if such a problem is not avoided, we also lose some of the true behavior of α's convergence to Kuzmin's Theorem if we do not include some values of k and n such that the bounding function captures the possibly large divergence caused by the $a_1 = 10^8$ problem. It is important that our estimations include some form of this behavior so that the obtained bounding function will preserve the inequality in Equation 3.4.83 for many k. If we choose k large and n large, we can capture some of this behavior but the large n will have enough of a dilutive effect so as to preclude impractical values of A and λ. This approach differs from letting $n = 1$ or $n = 2$ because there is absolutely *no* dilutive effect in this case for $a_1,\ a_2$ large.

Lastly, I chose to examine the values $k = 1, 2, 3, 4, 5$ because according to Kuzmin's Theorem, these values should occur with the highest frequency; therefore, any empirically estimated bounding function must bound Inequality 3.4.83 for these values of k. I chose the range $k_i = 95, \ldots, 100$ randomly, but wanted numbers that were large enough, such that their Kuzmin expected frequencies would be very small relative to $k = 1, 2, 3, 4, 5$s' frequencies. Also, I wanted numbers small enough to capture the possibly large divergence caused by the $a_1 = 10^8$ problem, but large enough so that the beginning n would dilute some of the erratic behavior caused by this problem, which was accomplished by considering $n = 7,070$ and $k = 100$. Therefore, this experiment should caputure the behavior of the bounding function $Ae^{-\lambda(n-1)}$ for most k (or for most k observed in the expansions our test set of α) over the range $n = [7,070\ ,\ 2,000,000]$.

For each value of n (use n_i to distinguish distinct values of n) and for each value of k, I calculated the maximum divergence from Kuzmin's Theorem over all the examined α (label this maximum $y(n,k) = y(n)$, where y is really only a function of n because we fix k to determine this maximum divergence over all α at each n_i). Then for each k, I plotted the maximum divergence against the different values of n_i and found a best fit exponential decay function $d(n)$, which yielded values for λ and A. However, the best fit function did not bound $y(n)$ for all n because it was a trend line. Therefore, I obtained a best fit function $h(n)$ for each k, such that $h(n) \geq y(n)$ for $7,070 \leq n \leq 2,000,000$. Motivated by Equation 3.4.83, for $k = k_0$, I assumed $h(n)$ is of the form:

$$h(n) = \frac{A}{k_0(k_0+1)} e^{-\lambda(n-1)}. \tag{3.5.93}$$

The equation for $h(n)$ invloves two unknowns, A and λ, so I subjected $h(n)$ to the following conditions:

$$\begin{aligned} h(n_{7,070}) &= y(n_{7,070}) = V_0 \\ and,\ \ h(n_m) &= y(n_m) = V_1 \\ where,\ \ m : |d(n_m) - y(n_m)| &> |d(n_i) - y_(n_i)| \end{aligned} \tag{3.5.94}$$

where the last inequality holds for all $i \in (7,070\ ,\ 2,000,000]$. We can now find closed form expressions for A and λ in terms of our data.

We assumed in Equation 3.5.93 that $h(n) = \frac{A}{k_0(k_0+1)}e^{-\lambda(n-1)}$, so letting $n_{7,070} = n_0$ and $\frac{A}{k_0(k_0+1)} = B_{k_0}$ we begin with:

$$\begin{aligned} h(n_0) = B_{k_0}e^{-\lambda(n_0-1)} &= V_0 \\ h(n_m) = B_{k_0}e^{-\lambda(n_m-1)} &= V_1 \end{aligned}$$

Now we just solve the simultaneous equations for B_{k_0} and λ:

$$\begin{aligned} B_{k_0} &= V_0 e^{\lambda(n_0-1)} \\ \lambda &= \frac{-1}{n_m-1}\ln\left(\frac{V_1}{B_{k_0}}\right) \\ \Rightarrow \lambda &= \frac{-1}{n_m-1}\ln\left(\frac{V_1}{V_0 e^{\lambda(n_0-1)}}\right) \\ \Rightarrow e^{\lambda} &= \left(\frac{V_1}{V_0}\right)^{\frac{-1}{n_m-1}}\left(e^{\frac{\lambda(n_0-1)}{n_m-1}}\right) \\ \Rightarrow e^{\lambda(\frac{n_m-n_0}{n_m-1})} &= \left(\frac{V_1}{V_0}\right)^{\frac{-1}{n_m-1}} \\ \Rightarrow \lambda &= \left(\frac{1}{n_m-n_0}\right)\ln\left(\frac{V_0}{V_1}\right) \end{aligned} \tag{3.5.95}$$

Then substituting $B_{k_0} = V_0 e^{\lambda(n_0-1)}$ from line 1 into the last line of Equation 3.5.95, we obtain an expression for B_{k_0}:

$$B_{k_0} = V_0\left(\frac{V_0}{V_1}\right)^{\frac{(n_0-1)}{n_m-n_0}} \tag{3.5.96}$$

We have thus outlined a method for determining the constants A and λ in Levy's error term for a given range of n beginning with n_0.

Given a numerical analysis of a set of $\alpha \in K$, there exists two options for choosing the optimal values of A and λ. Note, we obtain a different bounding function $h_{k_i}(n)$ for each tested value of k_i. Therefore, our first choice for the values of A and λ is A_{k_m} and λ_{k_m}, where k_m is chosen such that:

$$A_{k_m}e^{-\lambda_{k_m}(n-1)} > A_{k_i}e^{-\lambda_{k_i}(n-1)} \tag{3.5.97}$$

for all k_i and for all n. Choosing these constants will yield the optimal bounding function for the given test set of α.

The other choice for the values of the constants is A_{k_n} and λ_{k_r}, where $A_{k_n} > A_{k_i}$ and $\lambda_{k_r} < \lambda_{k_i}$ for all i. While these choices are not optimal for our data, the resulting bounding function:

$$A_{k_n} e^{-\lambda_{k_r}(n-1)} \tag{3.5.98}$$

will certainly be more robust than the bouding function in Equation 3.5.97, in terms of satisfying Inequality 3.4.83 for a larger set of untested α and for more values of k.

If the set of tested α is a subset of a special family F of numbers (e.g. cube roots, all α have the same Galois group, etc.), then we can expect untested $\alpha \in F$ to behave roughly similar to the tested set of α. Additionally, some values of k in the expansions of the untested $\alpha \in F$ will produce bounding functions larger/less than $A_{k_n} e^{-\lambda_{k_r}(n-1)}$. Thus, taking as our bounding function the one presented in Equation 3.5.98 will yield a bounding function better suited for $\alpha \in F$ than the function in Equation 3.5.97.

3.5.4 Results

Before we present the empirical results to the above outlined experiment, we ask what factors should govern the values of A and λ? We expect A's value will be determined by whatever k (over all α) and n exhibit the largest divergence from Kuzmin's expectation. Due to the large digit problem, we expect that for some tested $\alpha \in K$ the largest observed divergence will be for the case of n-small and k-large, namely $n = 7,070$ and $k = 100$. In other words, we expect that for some α, the frequency of $a_i = 100$ for $i \leq 7070$ will be significantly different than Kuzmin's predicted frequency (i.e. the event $a_i = 100$ should occur once for $i \leq 7,070$). However, because Kuzmin's predicted frequency for $a_i = 100$, given $i \leq 7,070$, is so small, any α that does not have exactly one occurrence of $a_i = 100$ for $i \leq 7,070$ will produce a large divergence value, in terms of percent difference from Kuzmin's expectation.

We cannot rely exclusively on the same intuition to determine the factors that should affect the value of λ. If the bounding function for k produces the smallest λ, then the convergence of $\mu(\alpha : a_n = k)$ to $\log_2(1 + \frac{1}{k(k+2)})$ is slower than other values of k over the range of $n \in [7,070\ ,\ 2,000,000]$. Recall after fixing k, $y(n)$ is defined as the maximum divergence (over all α and over all n_i) from Kuzmin's expectation of the frequency of $a_j = k$ given continued fraction expansions of length n_i, or formally:

$$y(n) = \max_{\alpha, n_i} \left| \mu(E \binom{n_i}{k}) - \frac{\ln \left\{ 1 + \frac{1}{k(k+2)} \right\}}{\ln 2} \right|. \qquad (3.5.99)$$

Furthermore, the main factor affecting the value of λ is the behavior of the "tail" of $y(n_i)$, where the tail is defined as the values of $y(n_i)$ for i near the end of the range of n, or in our case $i \approx 2,000,000$. λ's value is most affected when the tail of $y(n_i)$ is an increasing function, but for all other i in the range of n, $y(n_i)$ follows an exponential decay trend (we assumed that $y(n_i)$ and the bounding function decay exponentially). Note, that if range of n were extended far beyond $n = 2,000,000$ and $y(n_i)$ were an increasing function for i near $2,000,000$ but exhibited exponential decay behavior for all other i, then this non-exponential decay behavior would have only a minimal effect on λ's value.

Again, it is important to note that the constants A and λ must be recomputed using different values of n, k, and α for different applications. The intention of this experiment was to find a method for approximating A and λ for a given an application.

We now present the empirically determined values of A and λ for each case of k. The values of λ should be intepreted as $(reported\ \ value) \times 10^{-6}$ and the values of A have been adjusted by multiplying each experimentially determined constant B in Equation 3.5.96 by the factor $k(k+1)$. (see Appendix A for the backup data, graphs and summary pages):

k	A	λ
1	0.0221	1.464
2	0.0191	0.825
2	0.1197	1.645
4	0.1075	1.452
5	0.1232	1.826
96	1.4386	1.052
97	1.4404	1.219
98	2.7181	1.543
99	1.4395	1.105
100	2.8870	1.469

The first conclusion one can draw from the data is both A and λ seem to depend on k. We mentioned earlier that in Equations 3.3.80 and 3.4.83, A and λ were assumed to be absolute positive constants, but if one attempts to assign numerical values to either A or λ then these constants become functions of k, which depends α vis a vis $a_i(\alpha) = k$. Thus, we lose the power of both Kuzmin's and Levy's error functions being independent of α to the extent that $\alpha \in K$; Kuzmin's and Levy's error terms clearly depend on α if $\alpha \notin K$.

Because all the tested αs belong to the family of cube roots of primes, I will choose for the optimal bounding function the largest A_{k_m} and the smallest λ_{k_r}; thus the value of A is determined by $k_m = 100$, as expected, and the value of λ is determined by $k_r = 2$, as expected after observing that for $k = 2$, $y(n_i)$ is an increasing function for all $i \geq 1,750,000$. To correct this problem we could reperform the experiment with a longer range of n; if we examined longer ranges of n, then this non-exponential decaying tail would not have such a substantial effect on the empirically estimated value of λ. But herein lies the problem with estimating such constants.

The approximating function obtained is:

$$f(n) = \frac{2.8870}{k(k+1)} \times e^{-8.2476 \times 10^{-7} \times (n-1)} \tag{3.5.100}$$

We expect these choices for A and λ should hold for many values of k for many α that are cube roots of prime numbers, or at least for the cube roots tested in [Mi]. We expect this bound to hold for the set of α tested in [Mi] because our bounding function in Equation 3.5.100 was obtained from the α in this set with the largest divergence from Kuzmin's expectations. The consequence of choosing the constants in such a fashion is our bounding function is not the optimal bound for our data because of reasons discussed previously.

It is important to point out that we could take A to be arbitrarily large and λ to be arbitrarily small, such that the inequality in Equation 3.4.83 is satisfied for almost all k (almost all α) and for all n, but then we lose accuracy in our approximation of the difference in Equation 3.4.83. An interesting question is what are the minimum values of A and λ such that the inequality in Equation 3.4.83 is satisfied by a set of α with full measure?

We finally return to our question as to whether the inequality in Equation 3.4.83 can be bounded by a function $C \times e^{g(n)} < A \times e^{-\lambda(n-1)}$, where $g(n)$ is of a higher order than n. Over a certain range of n, we certainly can find such a function. Two ways of constructing this function are to allow the value of C to be very large, or to let $g(n) = -\lambda n^t$ and allow the value of λ to be very small; we can make t arbitrarily large by making λ arbitrarily small. But in general, the larger the value t the smaller the range of n, for which $C \times e^{g(n)}$ satisfies the inequality in Equation 3.4.83. However, there is no data or theory that suggests such a function could/will hold over a range of n as $n \to \infty$.

Chapter 4

Bounded Coefficients

4.1 Known Theory

We present the following theorem to illustrate a lower bound for how fast x converges to α vis a vis the growth rate of the denominators.

Theorem 4.1.1. *For any $k \geq 2$, we have:*

$$q_k \geq 2^{\frac{k-1}{2}} \quad [\textit{MT}],\ [\textit{Ki}]. \tag{4.1.1}$$

Proof: For $k \geq 2$, we have that $q_k = a_k q_{k-1} + q_{k-2} \geq q_{k-1} + q_{k-2} \geq 2q_{k-2}$ by Theorem 1.4.2. Repeating this inequality we arrive at $q_{2k} \geq 2^k q_0 = 2^k$ and $q_{2k+1} \geq 2^k q_1 \geq 2^k \Rightarrow q_k \geq 2^{\frac{k-1}{2}}$. □

An alternative proof is provided in [MT], where the authors use the recurrence relations of the Fibonacci sequence to bound q_k from below.

This theorem implies that the denominators of the convergents do not increase more slowly than the terms of a certain geometric series. Even if all $a_i = 1$ (as is the case with $\alpha = \frac{1+\sqrt{5}}{2}$, which is the slowest converging continued fraction), q_n still grows at a rate equal to a geometric progression, which implies a very fast convergence rate, in general. However, we will see in the next theorem that the denominators cannot grow faster than e^{Bn}, which will provide useful insight into analyzing the behavior of large valued coefficients (i.e. $a_i = k$ for k very large).

Theorem 4.1.2. *There exists a positive absolute constant B such that for sufficiently large n the inequality*

$$q_n = q_n(\alpha) < e^{Bn} \tag{4.1.2}$$

holds almost everywhere [Ki].

I will not provide the proof of this theorem because we only need the statement of this theorem to analyze the results presented in this chapter. However, Khintchine gives a slightly more explicit form for B in the proof, namely $B = A + \log 2$. It is important to note that for almost all $\alpha \in [0,1]$, we now have a lower and an upper bound for the growth of the denominators of the convergents. Both bounds are geometric progressions depending on an absolute constant. These bounds will imply certain bounds applicable to digit values.

In other words, for almost all α we have $2^{\frac{n-1}{2}} < q_n < e^{Bn}$. Taking the n^{th} roots of the inequality yields $a < q_n^{\frac{1}{n}} < e^B$ for *almost all numbers* $\alpha \in [0,1]$, where $a = \lim_{n\to\infty} \sqrt[n]{2^{\frac{n-1}{2}}} = \sqrt{2}$. In fact, Levy proved that there exists an absolute constant γ such that

$$\lim_{n\to\infty} \sqrt[n]{q_n} = \gamma, \tag{4.1.3}$$

where

$$\ln(\gamma) = \frac{\pi^2}{12 \ln 2}. \tag{4.1.4}$$

If we combine Theorems 4.1.2 and 2.4.1, we find one subset of the set of α not satisfing Equation 4.1.2 is a set of transcendental numbers that have digit values violating both Inequality 4.1.2 and the inequality of Louiville's Theorem.

We now present a general result that motivated the empirical investigation presented later in this chapter. While the theorem is self-explanatory, the results are profound.

Theorem 4.1.3. *The set of all numbers in the interval* $[0,1]$ *whose coefficients are bouded is of measure zero, or* $\mu(\alpha \in [0,1] : a_i \leq M \ \ \forall i) = 0$

Proof: This proof can be found in [MT]. Recall that each rank n interval is a subset of some rank $n-1$ interval, or $J_n \subset J_{n-1}$. Consider Equation 3.2.18, which gives $\frac{1}{3k^2} < \mu(E \begin{pmatrix} n \\ k \end{pmatrix})$, or the measure of the set of α, whose n^{th} coefficient is k, is greater than $\frac{1}{3k^2}$. Intuitively, "this results shows that in any arbitrary interval of rank $n-1$, that interval of rank n which is characterized by the value $a_n = k$ takes up a part of (at least) $\frac{1}{3k^2}$" [Ki]. In other words, the measure of the rank n interval characterized by $a_n = k$ will be at most:

$$\mu(J_n^k) < (1 - \frac{1}{3k^2})\mu(J_{n-1}). \tag{4.1.5}$$

Applying this inequality to the theorem at hand, since $a_i = k_i < M \ \forall i$, we have for all $k < M$, $(1 - \frac{1}{3k^2})\mu(J_{n-1}) < (1 - \frac{1}{3M^2})\mu(J_{n-1})$, so we consider:

$$\mu(J_1^M) \leq (1 - \frac{1}{3M^2})\mu(J_0), \tag{4.1.6}$$

where $J_0 = [0, 1]$ and repeated application of this inequality yields:

$$\mu(J_n^M) \leq (1 - \frac{1}{3M^2})^n \mu(J_0). \tag{4.1.7}$$

Since$(1 - \frac{1}{3M^2}) < 1$, as $n \to \infty$, $\mu(J_n^M) \to 0$. □

Combining Theorems 2.3.6 and 4.1.3, we conclude that for a sufficiently small fixed constant c a number α with bounded coefficients cannot be approximated by a rational number better than

$$\left|\alpha - \frac{p}{q}\right| < \frac{c}{q^2}. \tag{4.1.8}$$

But the previous theorem states that the set of such numbers has measure zero, so almost all numbers can be approximated by a rational number to a degree better than $\frac{c}{q^2}$ in Equation 4.1.8. Note, in light of Theorem 2.4.1, quadratic irrationals also cannot be approximated better than an order of $\frac{1}{q^2}$, but the set of quadratic irrationals has zero measure. See [MT] for more details.

The next theorem will lead to a nice result to be presented later in Proposition 4.2.1.

Theorem 4.1.4. *Let $\phi(n)$ be an arbitrary positive function of the positive integer n. If the series $\sum_{n=1}^{\infty} \frac{1}{f(n)}$ diverges, then the inequality*

$$a_n = a_n(\alpha) \geq \phi(n) \tag{4.1.9}$$

is satisfied an infinite number of times for almost all α. On the other hand, if the series $\sum_{n=1}^{\infty} \frac{1}{\phi(n)}$ converges, then the inequality is satisfied $\mathbf{at\ most}$ *a finite number of times for almost all α [Ki].*

Proof: Let us consider the first statement of the theorem. Let J_{n+m} be an interval of rank $n+m$, such that the continued fraction expansions of all $x \in J_{n+m}$ satisfy:

$$a_{m+i} < \phi(m + i), \tag{4.1.10}$$

where $i = (1, 2, 3, \ldots, n)$.

Using the same notation from Theorem 4.1.3 and recalling from Equation 3.2.16 that $\mu(J_{n+1}^k) > \frac{1}{3k^2}\mu(J_n)$, we conclude:

$$\begin{aligned} \mu\left(\sum_{k\geq M} J_{n+1}^k\right) &> \frac{1}{3}\mu(J_n)\sum_{k\geq M}\frac{1}{k^2} \\ > \frac{1}{3}\mu(J_n)\sum_{i=1}^{\infty}\frac{1}{(M+i)^2} &> \frac{1}{3}\mu(J_n)\int_{M+1}^{\infty}\frac{du}{u^2} \\ &= \frac{1}{3(M+1)}\mu(J_n) \end{aligned} \tag{4.1.11}$$

and since

$$\begin{aligned} \sum_{k=1}^{\infty} J_{n+1}^k &= J_n, \\ \Rightarrow \mu\left(\sum_{k<M} J_{n+1}^k\right) &< \left\{1 - \frac{1}{3(M+1)}\right\}\mu(J_n), \end{aligned} \tag{4.1.12}$$

where the first inequality in 4.1.11 is a result of Equation 3.2.16. The second inequality is a result of reindexing the sum in line 1 and subtracting $\frac{1}{M^2}$. The third inequality is a result of letting $u = M + i$, and noting that this intergral is a refinement of the sum on the left hand side. The first line of 4.1.12 is a restatement of Equation 3.2.13 and combining the results with Equation 3.2.16 yields the final inequality.

Thus, letting $M = \phi(m + n + 1)$ in the last line of Equation 4.1.12:

$$\mu\left(\sum_{k<\phi(m+n+1)} J_{m+n+1}^k\right) < \left\{1 - \frac{1}{3(1+\phi(m+n+1))}\right\}\mu(J_{m+n}). \tag{4.1.13}$$

Let us sum this inequality over all rank $m + n$ intervals, whose elements satisfy the condition in Equation 4.1.10, and denote this collection of rank $m+n$ intervals by $E_{m,n}$. We obtain:

$$\mu(E_{m,n+1}) < \left\{1 - \frac{1}{3(1+\phi(m+n+1))}\right\}\mu(E_{m,n}), \tag{4.1.14}$$

where we lose at least $\left(1-\frac{1}{3(1+\phi(m+n+1))}\right)$ part of $E_{m,n}$ because $a_{m+n+1} < \phi(m+n+1)$ for all $\alpha \in E_{m,n+1}$.

Let the assumption in the first statement of the theorem hold, that is the series $\sum_{n=1}^{\infty} \frac{1}{\phi(n)}$ diverges; then the series $\sum_{i=2}^{\infty} \frac{1}{3(1+\phi(m+i))}$ diverges for any constant m. As a result, we fix m and argue:

$$\lim_{n\to\infty} \prod_{i=2}^{n} \left\{1 - \frac{1}{3(1+\phi(m+i))}\right\} \to 0$$

$$\Rightarrow \lim_{n\to\infty} \mu(E_{m,n}) < \lim_{n\to\infty} \prod_{i=1}^{n} \left(1 - \frac{1}{3(1+\phi(m+i))}\right)\mu(E_{m,1}) \to 0$$

$$\Rightarrow \lim_{n\to\infty} \mu(E_{m,n}) = 0, \quad (4.1.15)$$

where the second line is the combination of line 1 and Equation 4.1.14. Thus, we have that for any m, $\mu(E_{m,n}) \to 0$ as $n \to \infty$.

For a given m, let E_m denote the set of all $\alpha \in [0,1]$, such that $a_{m+i} < \phi(m+i)$ for all $i \in N$, which implies this set of α is a subset of every set: $E_{m,1}, E_{m,2}, \ldots, E_{m,n}, \ldots$. Then from Equation 4.1.15, we have $\mu(E_m) = 0$.

Let $E_1 + E_2 + \ldots + E_m + \ldots = E$, then $\mu(E) = \mu(E_1 + E_2 + \ldots + E_m + \ldots) \leq \sum_{m=1}^{\infty} \mu(E_m) = 0$. Every α, such that $a_n = a_n(\alpha) \geq f(n)$ is satisfied only a finite number of times, belongs to one of the sets E_m for a sufficiently large m, but $\mu(E_m) = 0 \;\; \forall m$. The first assertion is proved.

Now assume the series $\sum_{n=1}^{\infty} \frac{1}{\phi(n)}$ conveges. Denote the embedded rank $n+1$ interval, such that $a_{n+1} = k$, by $J_{n+1}^{k} \subset J_n$. We know from Equation 3.2.16:

$$\mu(J_{n+1}^{k}) < \frac{2}{k^2}\mu(J_n), \quad (4.1.16)$$

which implies:

$$\mu\left(\sum_{k\geq\phi(n+1)} J_{n+1}^{(k)}\right) < 2\mu(J_n)\sum_{k\geq\phi(n+1)}\frac{1}{k^2}$$

$$\leq 2\mu(J_n)\sum_{i=0}^{\infty}\frac{1}{\{\phi(n+1)+i\}^2} < 2\mu(J_n)\Big(\frac{1}{\phi(n+1)}+\int\limits_{\phi(n+1)}^{\infty}\frac{du}{u^2}\Big)$$

$$= \frac{4\mu(J_n)}{\phi(n+1)}, \tag{4.1.17}$$

where the first inequality is obtained by summing over all $k \geq \phi(n+1)$ in Equation 4.1.16. The second inequality is obtained by letting $k = \phi(n+1)+i$ and recognizing that $k \geq \phi(n+1)\ \ \forall n$. The third inequality is obtained by letting $u = \phi(n+1)+i \Rightarrow\ du =\ di$, and from calculus $\int\limits_{\phi(n+1)}^{\infty}\frac{du}{u^2} < \sum\limits_{i=0}^{\infty}\frac{1}{\{\phi(n+1)+i\}^2}$, therefore, the additional term $\frac{1}{\phi(n+1)}$ is required to make the third inequality strict. The final equality is a result of evalutating the integral $\int\limits_{\phi(n+1)}^{\infty}\frac{du}{u^2} = 0 - \frac{-1}{\phi(n+1)} = \frac{1}{\phi(n+1)}$ and collecting the terms.

Let F_n be the set of $\alpha \in [0,1]$, such that $a_n \geq \phi(n)$, and then sum the inequality obtained in Equation 4.1.17 over all rank n intervals J_n. Thus, we conclude:

$$\mu(F_{n+1}) < \frac{4}{\phi(n+1)}. \tag{4.1.18}$$

Since $\sum \frac{1}{\phi(n+1)}$ converges, the sets $F_1, F_2, \ldots, F_n, \ldots$ form a convergent series, and by the n^{th} term test, $\mu(F_n) \to 0$ as $n \to \infty$. Therefore, if we allow F to be the set of all $\alpha \in [0,1]$, which belong to infinitely many F_n, then $\mu(F) = 0$.

Justifying this final step is an exercise in metric set theory: (I will follow the proof presented in [Ki]). For any m, the set F is contained in the set $\sum\limits_{n=m}^{\infty} F_n$, which implies that $\mu(F) < \sum\limits_{n=m}^{\infty}\mu(F_n)$. Taking m to be sufficiently large, we can make $\mu(F)$ arbitrarily small. By construction, F is the set of all numbers for which the condition in Equation 4.1.9 is satisfied infinitely often. □

4.2 Motivation

Kuzmin's Theorem holds for all numbers $\alpha \in [0,1]$ except for a set of zero measure. An implication of Kuzmin's Theorem is the $Prob(a_i = k) > 0$ for all $k \in N$. In other words, for k arbitrarily large and $\alpha \in K$, there is a positive probability that for some coefficient a_i in α's expansion, we have $a_i = k$.

On the other hand, if α_0 has a continued fraction expansion whose coefficients a_i are bounded by M, then $Prob(a_i = k) = 0$ for all $k > M$ and all i. As a result, $a_0 \in Z$, where $[0,1] \setminus K = Z$.

Let α have a continued fraction expansion whose coefficients are bounded by a monotonically increasing function $g(i)$ (i.e. $a_i = k_i \leq g(i) \ \ \forall i$). What is the slowest growing positive function $g(i)$ such that $\alpha \in K$? Clearly, we have:

$$\lim_{i\to\infty} g(i) = \infty \tag{4.2.19}$$

Consider the following argument: let $\alpha_0 = [0; a_1, a_2, \ldots] \in [0,1]$ be irrational. For ease of exposition, we ignore the error term in Kuzmin's Theorem, then Equation 3.5.85 gives:

$$Prob(a_i = k) = \log_2\left(1 + \frac{1}{k(k+2)}\right), \tag{4.2.20}$$

which gives the value of $Prob(a_i = k)$ for all $k \in N$. Therefore, if k occurs with probability $Prob(a_i = k)$, then we expect for some $i \leq \frac{1}{Prob(a_i=k)}$ that $a_i = k$. In general, the frequency of $a_i = k$ is given by $\frac{1}{Prob(a_i=k)}$, or the event $a_i = k$ occurs once in every $\frac{1}{Prob(a_i=k)}$ coefficients. For illustration purposes, consider $k = 1, 2, 3, 4$ substituted into Equation 4.2.20:

$$\begin{aligned} Prob(a_i = 1) &= 0.4150 \\ Prob(a_i = 2) &= 0.1699 \\ Prob(a_i = 3) &= 0.0931 \\ Prob(a_i = 4) &= 0.0589 \end{aligned} \tag{4.2.21}$$

These computations correspond to observing $a_i = 1$ about once in every $\frac{1}{.4150} \approx 2$ digits, $a_i = 2$ once in every $\frac{1}{.1699} \approx 6$ digits, $a_i = 3$ once in every $\frac{1}{.0931} \approx 11$ digits, and $a_i = 4$ once in every $\frac{1}{.0589} \approx 17$ digits.

Thus, suppose our bounding function $g(i)$ took the following values $g(1) = 1,\ g(2) = 1,\ g(3) = 2,\ g(4) = 2,\ g(5) = 2,\ g(6) = 2,\ g(7) = 3,\ g(8) =$

$3,\ g(9) = 3,\ g(10) = 3,\ g(11) = 3,\ g(12) = 4,\ g(13) = 4,\ g(14) = 4,\ g(15) = 4,\ g(16) = 4,\ g(17) = 4,\ g(18) = 5,\ \ldots$, or $g(i)$ grows with $\log_2(i)$. By choosing $g(i)$ as our bounding function we could construct an expansion, whose coefficients obey Kuzmin's probability distribution in the limit. Since this expansion would hold in the limit, writing down one of the infinitly possible expansions would be virtually impossible. The general form of such an expansion is to allow all digit values $k' < g(i)$ for all i to occur with Kuzmin's expected frequency, and the event $a_i = k_0$ must occur once for $\frac{1}{\log_2(1+\frac{1}{(k_0-1)((k_0-1)+2)})} < i < \frac{1}{\log_2(1+\frac{1}{k_0(k_0+2)})}$, and then k_0 becomes one of the k's. See Proposition 4.4.1 for a formal analysis of this function.

Could $g(i)$ be the minimal bounding function, such that the set of all $\alpha \in [0,1]$, whose coefficients are bounded by $g(i)$, has full measure and also obeys Kuzmin's Theorem? The answer is no.

We argued the function $g(n) = \log_2(n)$ is the slowest growing bounding function, such that a continued fraction expansion can obey Kuzmin's Theorem in the limit. In light of Theorem 4.1.4, because the sum $\sum_{n=1}^{\infty} \frac{1}{g(n)}$ diverges, the inequality $g(n) \geq a_n$ is satisfied for only a finite number of a_n (finitely often) for almost all α (i.e. a set of full measure with possibly the exception of a zero measure set). In other words, $\mu(\alpha \in [0,1] : a_n \leq g(n) = \log_2(n)$ infinitely often$) = 0$, where infinitely often means as $g(n)$ grows with n the inequality $a_n \leq g(n)$ is satisfied for an infinite number of a_n. Thus, we must consider a faster growing function $\phi(n)$ to ensure that $\mu(\alpha \in [0,1] : a_n \leq \phi(n)$ infinitely often$) = 1$.

Proposition 4.2.1. *For an arbitrarily small constant $\epsilon > 0$, the function*

$$\phi(n) = n^{1+\epsilon} \tag{4.2.22}$$

is a positive growing function of n, such that $\phi(n) \geq a_n$ holds for all but a finite number of a_n for almost all $\alpha \in [0,1]$ (denote this set of α by S), and such that Kuzmin's Theorem is satisfied by almost all $\alpha \in S$.

Proof: Let $\phi(n) = n^{1+\epsilon}$ and note

$$\sum_{n=1}^{\infty} \frac{1}{n^{1+\epsilon}} \tag{4.2.23}$$

converges for any $\epsilon > 0$. Thus, by our definition of F in the proof of Theorem 4.1.4, the inequality $\phi(n) = n^{1+\epsilon} \geq a_n$ is violated only finitely often for almost all $\alpha \in [0,1]$ (see the analysis following Equation 4.1.18).

Let $K \bigcap S = F$. Since $\mu(\alpha \in K) = 1 = \mu(\alpha \in S)$ we have $\mu(F) = 1$, which is the set of $\alpha \in [0, 1]$ that obeys Kuzmin's Theorem and satisfies the inequality of the proposition for all but possibly a finite number of digits. □

Extending the analysis of the proof, for almost all $\alpha \in K$ the inequality of the proposition is satisfied for all but possibly a finite number of digits, and almost all $\alpha \in S$ obey Kuzmin's Theorem. In light of our discussion preceeding this proposition, since $n^{1+\epsilon} > \log_2(n)$ for all n, the function $n^{1+\epsilon}$ grows sufficiently fast such that Kuzmin's probability law *can* be satisfied for almost all $\alpha \in S$. Kuzmin's probability law cannot be satisfied for all $\alpha \in S$ because those α whose coefficients are bounded by $h(n) < \log_2(n)$ are elements of S.

The only shortcoming of Proposition 4.2.1 is the relationship $\phi(n) \geq a_n$ is violated possibly a finite number of times for $\alpha \in S$ and almost all $\alpha \in K$. But does there exist a "slowest" growing function such that $\phi(n) \geq a_n$ is satisfied for all n for almost all $\alpha \in K$?

One approach to finding such a function is to analyze the set $\alpha \in F$. In theory, we want to add a constant $M(\alpha)$ to $n^{1+\epsilon}$ for each $\alpha \in F$, such that the inequality $n^{1+\epsilon} + M(\alpha) > a_n$ holds for all n for this α. If we could determine the value of $M(\alpha)$ for α, we can then try to find one value of M that works for all $\alpha \in F$ simultaneously; it would be $M = \max_{\alpha \in F} M(\alpha)$ if this limit did indeed exist. But since $\mu(\alpha \in F) = 1$, we cannot find easily M's value because the number of $\alpha \in F$ is uncountable. However, for a given $\alpha_0 \in F$, we theoretically could find a function of the form $\theta(n) = n^{1+\epsilon} + M$ such that $\theta(n) \geq a_n(\alpha_0)$ for all n. By Proposition 4.2.1, it is possible to find this function for all $\alpha \in F$ because $\phi(n) \geq a_n$ is violated only a finite number of times.

The implication of Theorem 4.1.4 and Proposition 4.2.1 is almost all numbers $\alpha \in [0, 1]$ have unbounded coefficients, which is easy to see since $\lim_{n\to\infty} n^{1+\epsilon} = \infty$. However, from Proposition 4.2.1 we conclude the digits in the expansions of almost all $\alpha \in [0, 1]$ cannot become unbounded too quickly, too often. Clearly, we could make the same statement for almost all $\alpha \in K$.

Therefore, except for a set of zero measure, all $\alpha \in K$ satisfy Proposition 4.2.1, and as a result, Z consists of all α, whose coefficients satisfy $a_n \leq log_2(n)$ except for possibly a finite number of digits.

By Proposition 4.2.1, if α's coefficients satisfy $a_n \leq n^{1+\epsilon}$ only finitely often, then Inequality 3.4.83 cannot be satisfied for *all* values k, which implies $\alpha \in Z$.

4.3 Results

Let L be an arbitrarily large number and $\alpha \in K$, then it is extremely difficult to test the strength of Kuzmin's Theorem for $k \geq L$ because the probability that $a_i = L$ is very small. Consider the following: Kuzmin predicts that $Prob(a_i = L) = \log_2(1 + \frac{1}{(L)(L+2)}) < \epsilon$, where ϵ is arbitrarily small for sufficiently large L. Thus, we should observe $a_i = L$ for some $i < \frac{1}{\epsilon}$, but for L sufficiently large and ϵ sufficiently small, computers will not distinguish L from ∞ or ϵ from 0.

If $a_i \neq L$ for some $i < A$, where A is regarded as the maximum number of coefficients that can be computed within "reasonable" time by a computer and a computer can distinguish A from ∞, then we could never determine whether $a_i = L$ occurs, in the limit, with a frequency commensurate to Kuzmin's expecation. This problem becomes especially difficult to circumvent if $\frac{1}{\epsilon} > A$. Thus, the problem with testing Kuzmin's Theorem for $a_i = L$ is we cannot compute enough coefficients to verify Kuzmin's probability law (Equation 3.5.85); however, because of the theoretical results in Chapter 3, we must have faith and assume Kuzmin's Theorem holds for all values of L with an error term that is a function of both L and n.

Reiterating the logic mentioned above: if a continued fraction expansion of an irrational number is bounded, then this expansion cannot possibly obey Kuzmin's Theorem (Equation 3.5.85) for all values of k. As a result, for $\alpha \in K$, we expect to observe arbitrarily large coefficient values somewhere in its expansion. In fact, for $\alpha \in K$ and $k \in N$, the probability of never observing the event $a_i = k$ is:

$$\lim_{n \to \infty} (1 - Prob(a_i = k))^n = 0 \tag{4.3.24}$$

Based on research performed last year by Princeton University undergraduates under the tutelage of professors Ramin Takloo-Bighash and Steven Miller, Kuzmin's Theorem appears to hold in many different forms for the continued fraction expansions of prime roots of prime numbers (see [MT] for a summary of some of the results). The behavior of these expansions truncated after the first $500,000$ digits comported with Kuzmin's Theorem.

Motivated by these results, we examined the the first five prime roots of the first 117 prime numbers and the same roots of the first 100 primes greater than 10^8, denote by T this set of test α. After having Mathematica compute the first 10^6 coefficients in the continued fraction expansion of each $\alpha \in T$, I determined the maximum digit value and its position, as well as the values of a_{m-1} and a_{m+1}. Here, I will present the most astonishing results and a segway into the theoretical explanations governing our results.

The average maximum coefficient value for $\alpha \in T$ was 16,058,523, which Kuzmin predicts should occur with a probability of 5.45×10^{-15} implying that $a_i = 16,058,523$ should occur once every 1.83×10^{14} coefficients; however the average position of the maximum digit (a_m) was $m = 501,381$, and we note $501,381 \ll 1.83 \times 10^{14}$. The largest observed coefficient value was in the expansion of $619^{\frac{1}{3}}$: the $326,959^{th}$ digit's value was $2,625,830,672$, which occurs with probability 2.09×10^{-19}. The fact that such large coefficient values consistently occurr very early in the continued fraction expansions of $\alpha \in T$ may imply the continued fraction expansions of our test α generally converge to their true value faster than other $\alpha \in K$. I now offer a summary table of the data sorted by roots; note the "Min-Max Digit" is the lowest maximum digit value of all the tested α for each root, or for each root the Min-Max Digit $= \min_{\alpha \in T}$(Max. Digit(α)): (see Appendix A for full results)

Category	$\sqrt[3]{\alpha}$	$\sqrt[5]{\alpha}$	$\sqrt[7]{\alpha}$	$\sqrt[11]{\alpha}$	$\sqrt[13]{\alpha}$
Max Digit	2.63×10^9	2.99×10^8	2.01×10^9	1.79×10^9	1.21×10^9
Pos. of Max	3.27×10^5	6.74×10^5	9.67×10^5	2.22×10^5	7.84×10^5
Min-Max Dig.	2.09×10^5	1.91×10^5	2.47×10^5	2.57×10^5	1.94×10^5
Pos. of Min-Max	6.47×10^5	6.21×10^4	6.42×10^5	5.01×10^5	1.35×10^5
Avg. Max	2.17×10^7	7.24×10^6	1.89×10^7	1.30×10^7	1.94×10^7
Prob. of Max	2.09×10^{-19}	1.62×10^{-17}	3.57×10^{-19}	4.51×10^{-19}	9.78×10^{-19}
Prob. of Min-Max	3.30×10^{-11}	3.95×10^{-11}	2.36×10^{-11}	2.18×10^{-11}	3.84×10^{-11}

While the results seem to vary across the different prime roots, Kuzmin's Theorem suggests if we were to consider a sufficiently large n, all the maximum values should be relatively similar. A possible explanation for the 'early' occurrence of these large valued digits is the event $a_m = k_m$ (where m is defined above) will not reoccur for all $i \leq \frac{1}{Prob(a_i=k_m)}$. Because Kuzmin's Theorem holds in the limit, if we observe r occurrences of the event $a_i = k_m$ in the first $\frac{1}{Prob(a_i=k_m)} = i$ coefficients, then we expect the continued fraction expansion to possess $r-1$ increments of coefficients of length $\frac{1}{Prob(a_i=k_m)}$, such that $a_i \neq k_m$ for all i in these increments. However, due to computational limitations, we cannot conclude with full certainty that the expansions of our test α behave in such a manner; therefore, we cannot conclude that our data do or do not fall perfectly in line with Kuzmin's Theorem.

For $\alpha \in T$, we set out to show empirically the digits of a continued fraction expansion can assume arbitrarily large values, and our data do not reflect a systematic bound for the values of the coefficients, which may imply the coefficients of $\alpha \in T$ are unbounded. Observing very large digit values so early in the expansion lends more evidence to the theory that the coefficient values in the continued fraction expansions for almost all $\alpha \in [0, 1]$ can be arbitrarily large, which was the expected result.

Because our test α were also elements of K, we expected for at least one $\alpha_0 \in T$, the maximum valued coefficient $a_m(\alpha_0) = L$, where Kuzmin expects the event $a_i = L$ to occur once every 10^6 coefficients. However, we showed empirically the minimum-maximum valued coefficient had value $L = 191, 228 \Rightarrow Prob(a_i = L) = 3.95 \times 10^{-11}$, which means the event $a_i = 191, 228$ should occur approximately once every 25 billion coefficients.

Do prime roots of prime numbers consistently disobey Kuzmin's Theorem regarding the occurrence of large valued digits? Could the occurrence of such large values suggest that continued fractions converge faster to their true value α than Kuzmin predicts? Or, could these large values indicate a faster convergence rate for the first n coefficients of a given continued fraction than for the tail of its expansion? Does there exist a correlation between the position of the maximum digit and the value of the maximum digit? Does there exist a correlation between the value of the coefficient before/after the maximum valued coefficient and the value of the maximum coefficient? Theoretical answers to these questions will be provided below. The empirical results on the correlation questions can be found in Appendix A.

4.4 Possible Theoretical Explanations of the Results

From Proposition 4.2.1, let

$$\phi(n) = n^{1+\epsilon}, \tag{4.4.25}$$

where $\epsilon > 0$ is arbitrarily small. Intuitively, letting $\phi(n) = n^{1+\epsilon}$ means that for all but possibly a finite number of coefficients, the value (k_i) of the coefficient a_i will not exceed its position (i) almost everywhere. Comparing this theoretical expectation to our results by replacing n with m in Equation 4.4.25, we note for almost every $\alpha \in T$ we observed $a_m > m^{1+\epsilon}$, where we recall m denotes the position of the maximum valued coefficient in the expansion of α to 10^6 coefficients and $\epsilon > 0$ is an arbitrarily small constant; but the inequality $a_n \leq n^{1+\epsilon}$ should hold for only a finite number of a_n according to the Proposition. Thus, while our empirical results seem extraordinary, we can write them off as one of the finite violations of the inequality $a_n \leq \phi(n)$, but some of our results do indeed comport with the Proposition.

By examining the zero measure set of α for which Kuzmin's Theorem does not hold (denoted by Z), we can determine if the behavior of our extraordinary $\alpha \in T$ warrants classification of these α as elements of Z despite their Kuzmin-like behavior shown in [Mi].

4.4.1 Kuzmin's Measure Zero Set

The zero measure set Z, for which Kuzmin's Theorem does not hold, has not been described in an explicit form. However, based on empirical studies as well as theoretical arguments, some of the subsets comprising Z can be described. It is important to note that while a set $Z_i \subset Z$, there may be some $\alpha \in Z_i$, for which Kuzmin's Theorem does indeed hold for most values of k. Therefore, the task at hand is to locate those sets $Z_i \subset Z$ with the property that for all $\alpha_0 \in Z_i$, there exists k such that $Prob(a_i(\alpha_0) = k) = 0$, or such that $Prob(a_i(\alpha_0) = k) \neq \log_2(1 + \frac{1}{k(k+2)})$.

The first zero measure subset of Z is the set Q, or the rational numbers. Since rational numbers have finite continued fraction expansion, the expansion terminates, and we can actually determine the distribution of each k for each $\alpha \in Q$. More importantly, we can determine the maximum valued coefficient $a_m(\alpha) = M$ in the expansion of each $\alpha \in Q$; thus, for each $\alpha \in Q$ there exists $a_m(\alpha) = M$, such that $a_i(\alpha) \leq M \;\; \forall i$ and $Prob(a_i(\alpha) = k > M) = 0 \Rightarrow Q \subset Z$. However, it may be the case that some $\alpha \in Q$ appear to obey Kuzmin's Theorem for a finite number of values k throughout their continued fraction expansions; but these α cannot obey Kuzmin's Theorem for all k.

The second zero measure subset (I) of Z is the set of quadratic irrationals. This set has zero measure because it is a countable set. The quadratic irrationals are included in Z because their continued fraction expansion is periodic (see [Ki]); therefore for each $\alpha \in I$ there exists a maximum value coefficient $a_m(\alpha) = M$ (i.e. $a_i(\alpha) \leq M \;\; \forall i$) and $Prob(a_i(\alpha) = k > M) = 0 \Rightarrow I \subset Z$. Included in the quadratic irrationals are the golden ratio $\frac{1+\sqrt{5}}{2}$, the set of α obtained in [Fi], and all α with continued fraction expansions of the form $[k, k, \ldots, k]$, all of which are zero measure sets because the number of elements in each set is countable.

By a similar argument given for the rationals, we conclude that the third zero measure subset of Z is the set (B) of α with bounded coefficients. No $\alpha \in B$ can satisfy Kuzmin's Theorem for all k because to each $\alpha \in B$ there corresponds a number M, such that $a_i(\alpha) < M$ for all i; as a result, for each $\alpha \in B$, we have $Prob(a_i(\alpha) = k > M) = 0 \Rightarrow B$ is not a subset of K. We know from Theorem 4.1.3 that $\mu(B) = 0$, and therefore $B \subset Z$.

Numerical tests that I conducted suggest $e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots]$ and $e^2 = [7; 2, 1, 1, 3, 18, 5, 1, 1, 6, 30, 8, 1, 1, 9, 42, 11, \ldots]$ are also elements of Z. Also, Lang [La] showed that certain rational functions of e are not in Kuzmin's set, and Dimofte [Di] showed empirically that any "linearly periodic continued fraction retains its linear periodicity under a rational scaling or shift," which implies this set of α are not in Kuzmin's set. Additionally, we can include an uncountable collection of irrational and transcendental numbers that do not obey Kuzmin's Theorem as elements of Z without violating $\mu(Z) = 0$.

Finally, we are ready to add two additional subsets of Z.

Proposition 4.4.1. *Let $\gamma(i)$ be a positive growing function of i such that for all positive integers i we have $\gamma(i) < g(i)$, where $g(i)$ grows with $\frac{1}{\log_2(1+\frac{1}{k(k+2)})}$. Let G be the set of $\alpha \in [0, 1]$ whose coefficients satisfy $\gamma(i) \leq a_i(\alpha)$ only finitely often. Then $G \subset Z$ and $\mu(G) = 0$.*

Proof: Let $\gamma(i) < g(i)$ for all i. Kuzmin's Theorem (Equation 3.3.80) implies for $\alpha \in K$ we should observe $a_i(\alpha) = k$ for some $1 \leq i \leq \frac{1}{\log_2(1+\frac{1}{k(k+2)})}$. However, suppose that the digits $a_i(\alpha)$ were bounded by the function $g(i) = \frac{1}{\log_2(1+\frac{1}{i(i+2)})}$, which grows like $\log_2(i)$. Then as $\alpha \in K$ we expect to see the event $a_i = k$ once for some $g(i) \approx k - 1 < i \leq g(i) \approx k$, where we use "$\approx$" because $g(i) \notin Z$ for all i. If the digits of an expansion are bounded by $\gamma(i)$, then we would expect to see the event $a_i = k$ once for some $\gamma(i) \approx k-1 < i \leq \gamma(i) \approx k$. But since we we assumed $\gamma(i) < g(i)$, we know that $\gamma(i_0) \approx k = g(i_1) \approx k$ implies $i_0 > i_1$.

Let N be an arbitrarily large integer, then there exists k_0 such that Kuzmin expects the event $a_i = k_0$ to occur once for $i \leq N$, and this expected frequency could be satisfied by the expansion whose coefficients are bounded by $g(i)$ if $a_i = k_0$ for some $g(i) \approx k_0 - 1 < i \leq g(i) \approx k_0$; by construction, the i corresponding to $g(i) \approx k_0$ is $i = N$. Using the same k_0, there exists $i = N'$ such that $\gamma(N') \approx k_0$, but since $\gamma(i) < g(i)$ we have $N' > N$ and the expansion whose coefficients are bounded by $\gamma(i)$ cannot obey Kuzmin's expected frequencies for digit value k_0.

If we let $N \to \infty$, we observe such expansions cannot even satisfy Kuzmin's expected frequencies in the limit for all k because we can continue choosing N

arbitrarily large, finding a corresponding k_0, and arguing as above. Furthermore, if we apply the argument above for infinitely many N, then not even the finite violations of the inequality $\gamma > a_n(\alpha)$ can warrant classifying α with such expansions as elements of K. Finally, since $\sum_{n=1}^{\infty} \frac{1}{\log_2(n)}$ we have $\sum_{n=1}^{\infty} \frac{1}{\gamma(n)}$ diverges, and by Theorem 4.1.4 we conclude $\mu(G) = 0$. □

Can we strengthen the previous proposition by finding $G^{'} \subset Z$, such that for every $\alpha \in G^{'}$ we have $a_n \geq \gamma(n)$ infinitely often? The answer is no by Theorem 4.1.4 and Proposition 4.2.1.

Before we describe the final set $Z_i \subset Z$, we need to present an amazing Theorem proved by Khintchine. I will not present the proof, but we must understand that the proof of this theorem relies upon and is derived from Kuzmin's Theorem.

Theorem 4.4.2. *Let $f(r)$ be a non-negative function of the positive integer r. Further let positive constants C and δ exist such that:*

$$f(r) < Cr^{\frac{1}{2}-\delta}, \tag{4.4.26}$$

for $r = 1, 2, 3, \ldots$. Then for all numbers $\alpha \in (0, 1)$, with the exception of those of a set of measure zero, we have:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(a_k) = \sum_{r=1}^{\infty} f(r) \frac{\log\left(1 + \frac{1}{r(r+2)}\right)}{\log 2}, \tag{4.4.27}$$

where the convergence of the series follows from the conditions imposed on $f(r)$. [Ki]

Let us assume that this theorem is true and let $f(r) = \log r$ for $r = 1, 2, 3, \ldots$, then the condition set forth in Equation 4.4.26 is satisfied. Thus, as $n \to \infty$ the following relation holds almost everywhere:

$$\frac{1}{n} \sum_{k=1}^{n} \log(a_k) \to \sum_{r=1}^{\infty} \log(r) \frac{\log\left(1 + \frac{1}{r(r+2)}\right)}{\log 2}. \tag{4.4.28}$$

If we raise each side of this relation as a power of e and use the fact that $C\log(a) = \log(a^C)$, then we have:

$$\sqrt[n]{a_1 a_2 \cdots a_n} \to \prod_{r=1}^{\infty} \left\{1 + \frac{1}{r(r+2)}\right\}^{\frac{\log(r)}{\log 2}}. \tag{4.4.29}$$

We conclude for almost all $\alpha \in (0,1)$ and as $n \to \infty$, the geometric mean of the first n coefficients tends to an absolute constant given by:

$$\prod_{r=1}^{\infty} \left\{ 1 + \frac{1}{r(r+2)} \right\}^{\frac{\log(r)}{\log 2}} = 2.68545. \qquad (4.4.30)$$

While we recognize that this mean is commonly referred to as Khintchine's constant, we will also call this constant Kuzmin's expected geometric mean.

For all $\alpha \in K$, Kuzmin's Theorem gives a probability distribution of the digit values k, so the area under this distribution should be equal to 1. However, we cannot compute the expected value of k because the series $\sum_{k=1}^{\infty} k \log_2 \left(1 + \frac{1}{k(k+2)}\right)$ diverges. Since we cannot find the expected value of k, we should next examine what conclusions we can make regarding $\prod_{i=1}^{n} a_i$ for a typical $\alpha \in K$.

In the proof of Theorem 4.1.2 (see [Ki]), one would have reasoned if $g = e^{An}$ then $\mu(E_n(g) = \{\alpha \in (0,1) : a_1 a_2 \cdots a_n \geq g\}) = 0$. This inequality implies for sufficiently large n and for almost all α we have:

$$a_1 a_2 \cdots a_n < e^{An} \qquad (4.4.31)$$

By manipulating Kuzmin's Theorem and using Theorem 4.4.2 and Equations 4.4.30 and 4.4.31, we have for $\alpha \in K$: [Ki]

$$\prod_{k=1}^{\infty} \left\{ 1 + \frac{1}{k(k+2)} \right\}^{\frac{\log(k)}{\log 2}} = 2.68545. \qquad (4.4.32)$$

Therefore, Equation 4.4.30 is particularly useful in expressing the expected value/behavior of the digit values in an expansion of a typical $\alpha \in K$.

We are now ready to present our final subset of Z.

Proposition 4.4.3. *Let T denote the set consisting of all α, such that for an absolute constant $A > 1$ and a sufficiently large n, we have:*

$$q_n = q_n(\alpha) < 2^n e^{An}. \qquad (4.4.33)$$

Then the set $\Big((0,1) \setminus T\Big) = F \subset Z$, where Z is the zero measure set for which Kuzmin's Theorem does not hold.

Proof: By Theorem 4.1.2, as $n \to \infty$ the inequality $q_n < 2^n e^{An}$ is satisfied for almost all $\alpha \in (0,1)$ (again, this set of α is denoted by T). By Theorem 4.4.2, as $n \to \infty$ all $\alpha \in K$ satisfy the product relationship in Equation 4.4.30. Combining these two results yields $\mu(T \bigcap K) = 1$ because both sets T and K have full measure. However, we will show there exists a subset $T' \subset T$ such that $T' \not\subseteq K$ and $\mu(T') = 0$. If T' does exist, then $K \subset T$ and not $T \subset K$.

As a consequence Theorem 4.4.2, we have $\prod_{k=1}^{n} a_k = (2.68545)^n$ for all $\alpha \in K$ as $n \to \infty$, where a_k is the k^{th} digit of the continued fraction expansion of α. Thus, by Theorems 4.1.1 and 4.1.2 we know there exist two sets $T_1, T_2 \subset T$ but that are not subsets of K.

Let T_1 be the set of α whose coefficient product converges to a number less than $(2.68545)^n$ as $n \to \infty$ (for example consider $\alpha\frac{1+\sqrt{5}}{2}$). Clearly, $T_1 \subset T$ but as a result of Theorem 4.4.2 (in particular Equation 4.4.30) we have $T_1 \bigcap K = \emptyset$ so $T_1 \subset Z$.

Let T_2 be the set of α whose coefficient product converges to some number between $(2.68545)^n$ and e^{An}, where $A > 1$. Then $T_2 \subset T$, but $T_2 \bigcap K = \emptyset$ by Theorem 4.4.2, so $T_2 \subset Z$. Both T_1, T_2 have zero measure because $\mu(K) = 1$.

Since $\mu(K) = \mu(T) = 1$ and there exists $T_1, T_2 \subset T$ but that are not subsets of K, we have $K \subseteq T$.

Let $F = [0,1] \setminus T$, which implies the coefficient product of every $\alpha \in F$ converges to a number greater than e^{An}. By Theorem 4.1.2, $\mu(F) = 0$ and by Theorem 4.4.2 (in particular Equation 4.4.30) we have $F \not\subset K$. Thus $F \subset Z$.

Collecting our results, we conclude there exists a zero measure set $Z' = T_1 \bigcup T_2 \bigcup F$ such that $Z' \subset Z$.□

It is important to point out Z' does not necessarily correspond to the whole set Z. Thus, the result of the previous proposition is to give another characterization of the elements of K. Furthermore, Z' provides us a way to test empirically if $\alpha \in K$ simply by examining α's coefficient product. Although, in practice, we convert this product to a sum by using logarithms. Now we can use Equation 4.4.30 combined with the previous proposition to ascertain if our previous data on large valued coefficients are too extraordinary to be classified as Kuzmin numbers. Intuitively, we expect the α with the largest maximum coefficient values will have coefficient products greater than Khintchine's constant, which would imply these expansions converge faster in the first 10^6 coefficients to their true values than Khintchine predicts in Equation 4.4.30.

To put this convergence hypothesis to the test, I considered the five α with the largest maximum coefficient values in the first 10^6 coefficients, and the five α with the smallest maximum coefficient values over the same range. I then computed the product of the first 10^6 coefficients for these ten α:

$$\prod_{i=1}^{10^6} a_i. \tag{4.4.34}$$

I computed the absolute constant in Equation 4.4.30 by evaluating the product from $r = 1$ to $r = 10^6$ and from $r = 1$ to $r = 10^8$, both of which yielded the constant 2.68545 (in actually implementing the program, I converted the product into a sum using properties of logarithms). Later I confirmed this result with Mathematica's table of constants. Thus, we can test Kuzmin's expected geometric mean versus the observed geometric mean of the coefficients in the expansions of the ten suggested α. The coefficient products for these ten α are presented below, along with Kuzmin's expected geometric mean raised to the $10^{6^{th}}$ power.

$\underline{Number}$	$\underline{MaxValue}$	$\underline{Product}$	$\underline{Act. - Kuzmin}$	$\underline{\frac{Act.-Kuz.}{Kuz}}$
$\underline{KuzminExpected}$	N/A	1.18×10^{429017}	0	0
$\underline{SmallerMaximums}$				
$\sqrt[5]{149}$	191228	8.60×10^{429851}	$\approx 8.60 \times 10^{429851}$	7.29×10^{834}
$\sqrt[3]{467}$	209076	7.88×10^{429253}	$\approx 7.88 \times 10^{429253}$	6.68×10^{236}
$\sqrt[13]{613}$	193849	4.23×10^{429279}	$\approx 4.23 \times 10^{429279}$	3.58×10^{262}
$\sqrt[7]{10000439}$	247303	2.90×10^{429880}	$\approx 2.90 \times 10^{429880}$	2.46×10^{863}
$\sqrt[13]{10001461}$	216987	1.19×10^{428903}	$\approx -1.18 \times 10^{429017}$	≈ -1
$\underline{GreaterMaximums}$				
$\sqrt[13]{11}$	1214823489	5.73×10^{428485}	$\approx -1.18 \times 10^{429017}$	≈ -1
$\sqrt[11]{337}$	1789321825	8.67×10^{428560}	$\approx -1.28 \times 10^{429017}$	≈ -1
$\sqrt[7]{389}$	2009559864	1.59×10^{429429}	$\approx 1.59 \times 10^{429429}$	1.35×10^{412}
$\sqrt[3]{619}$	2625830672	5.21×10^{428858}	$\approx -1.28 \times 10^{429017}$	≈ -1
$\sqrt[13]{10001207}$	809115083	8.66×10^{429083}	$\approx 8.66 \times 10^{429083}$	7.24×10^{66}

Two phenomena within these results are worth noting. First, six of the ten coefficient products examined were significantly larger than Kuzmin expected in both absolute and relative terms, which could be explained by considering the possibility that continued fraction expansions converge faster to their actual underlying number than Kuzmin expects in Equation 4.4.30. Perhaps, in the limit (i.e. $n \gg 10^6$), the empirically observed convergence rates eventually fall in line with the theoretical convergence rates expected by Kuzmin, but this fact can be shown best through more numerical experiments. A final possibility is the six α, whose coefficient products were larger than Kuzmin's expectation, do not belong to our Kuzmin set K. However, given the analysis in [Mi], it is difficult to believe this explanation accurately describes the behavior of the data.

Secondly, four of the five α's with the smallest maximum coefficients exhibited coefficient product values that were significantly larger than the value predicted by Kuzmin. On the other hand, only two of the five α's with the greatest maximum coefficients had coefficient product values significantly greater than the value expected by Kuzmin. This result is quite astonishing and suggests the following question: Are continued fraction expansions with an early occurrence of a large coefficient value balanced out more than the expansions with a smaller maximum coefficient value by a greater number of low valued coefficients (i.e. $k = 1, 2$)?

The above data may further imply convergence to α is slower for those α with an extremely large coefficient value in the first 10^6 coefficients than for those α with smaller maximum coefficient values over the same range. However, it would still seem peculiar that continued fractions with the smallest maximum coefficients would converge significantly faster than those with the greatest maximum coefficients. Perhaps, we can best explain this observation by concluding the set of α, whose continued fraction expansions have a very large coefficient value early in the expansion, is not approximated as well by rational numbers as those α, whose continued fraction expansions have much smaller maximum coefficient values over the same range (i.e. $n = 10^6$).

If we exclude all $\alpha \in Z$, we question if irrationals whose coefficient values are unbounded but remain relatively small (i.e. small relative to other α being examined) actually converge faster to their true values than those irrationals whose coefficient values are unbounded and actually assume arbitrarily large values, given an analysis of n coefficients. If so, this could imply a subclassification of the irrationals into those irrationals that can be represented well by a fraction and those irrationals for which it is harder to represent as a rational number.

Linking this data analysis to our theory of unbounded coefficients is a bit difficult due to the lack of sufficient empirical research and theoretical constructs. Some of the general conclusions we can draw from these data lead to more very interesting topics of potential research.

To test the hypothesis that continued fraction expansions, with an early occurrence of a very large maximum coefficient value, are balanced out more than the expansions, with smaller maximum coefficient values, by a greater number of low valued coefficients (i.e. $k = 1, 2$) (hereinafter "the balancing hypothesis"), we present the following experiment and results. For both groups of α tested in the previous experiment, we will determine if the number of coefficients with values $k = 1, 2, 3,$ or 4 significantly diverges from Kuzmin's expectation for these values of k, see Equation 3.5.85:

$\underline{Number}$	$\underline{1's}$	$\underline{2's}$	$\underline{3's}$	$\underline{4's}$	$\underline{Total\ Low\ Digits}$
$\underline{Kuzmin\ Expected}$	415037	169925	93109	58894	736965
$\underline{Smaller\ Maximums}$					
$\sqrt[5]{149}$	413998	170019	93285	59044	736346
$\sqrt[3]{467}$	414896	170057	93127	58995	737075
$\sqrt[13]{613}$	414216	170856	92950	58683	736705
$\sqrt[7]{10000439}$	415125	169654	92893	58305	735977
$\sqrt[13]{10001461}$	414672	170571	93194	59168	737605
$\underline{Larger\ Maximums}$					
$\sqrt[13]{11}$	415489	169796	93408	58839	737532
$\sqrt[11]{337}$	415310	170015	92811	58843	736979
$\sqrt[7]{389}$	415126	169549	92653	59264	736592
$\sqrt[3]{619}$	415352	170358	93073	58508	737291
$\sqrt[13]{10001207}$	415710	168921	93013	59074	736718

The following conclusions hold for expansions of length $n = 10^6$. We do not observe a significant difference between the two sets of α in the number of total low valued coefficients; however, there is a substantial discrepancy between the two sets of α with regard to the number of observed $1's$. Four of the five α in the smaller maximum group have fewer $1's$ in their expansion than Kuzmin's expects for a typical $\alpha \in K$. Unfortunately, the α with more 1's than Kuzmin expects is not the α, whose coefficient product is smaller than Kuzmin's expected geometric mean (determined in the last experiment).

On the other hand, every α in the greater maximum group has more $1's$ in their expansions than Kuzmin predicts. In fact, the α in the smaller maximum group with the most 1's still has a fewer 1's in its expansion than in the expansion of the α in the greater maximum group with the least number of 1's.

If we compare the average number of $1's$ of each group, we find the smaller maximum α have on average $414,581$ coefficients with the value 1, and the larger maximum α have on average $415,397$, which corresponds to a difference of 816. In other words, in the expansion of our test α to 10^6 coefficients, we observed on average 816 more $1's$ in the greater maximum group than in the smaller maximum group. Since the larger maximum set of α consistently has more 1's than the smaller maximum set of α, we could infer the α in the larger maximum set converge to their true values slower than the α in the smaller maximum set. Therefore, we may be able to justify empirically our proposed subclassification of the irrationals: the set of larger maximum α are 'less rational' than the set of smaller maximum α.

Two possible explanations could justify our results:

First, perhaps the difference is explained by Levy's error term. We cannot apply the error function found in Equation 3.5.100 becauses the ranges of n are significantly different. The value of λ in Equation 3.5.100:

$$f(n) = \frac{2.8870}{k(k+1)} \times e^{-8.2476 \times 10^{-7} \times (n-1)} \tag{4.4.35}$$

was governed by the behavior of the tails (i.e. $1,800,000 \leq n \leq 2,000,000$) of the different cases of k. Therefore, our bounding function will not apply to our present data.

We can calculate the error function's required value in order to explain the disparate number of 1's by Levy's expected error. All cases of the different number of coefficients with value 1 must be explained by our error function, which implies our limiting cases are $\sqrt[5]{149}$ with $413,998$ observed 1's and $\sqrt[13]{10001207}$ with $415,710$ observed 1's. In other words:

$$[413,998; 415,710] \subset 415,037 \pm \ (error\ function\ value\ \times 10^6), \tag{4.4.36}$$

which gives an error function value of $.001037$. This value seems reasonable given 0.277 is the value of the error function in Equation 3.5.100 evaluated at $k = 1$ and $n = 2 \times 10^6$. Therefore, our set of test α appears to obey Kuzmin's Theorem with regard to the expected number of coefficients with values $k = 1\ or\ 2$, but the disparate products are still unexplained.

If in fact Levy's error function could not explain our results, then consider the possibility that the balancing hypothesis is true. Then if we see on average 816 less $1's$ in the expansion of the smaller maximum α, we must conclude that these 816 coefficient values are at least 2, where this average holds for an expansion to 10^6 coefficients. Therefore, the smaller maximum α have an additional factor in their coefficient product of at least $2^{816} = 4.37 \times 10^{245}$, which is substantial given this is a lower bound of the additional factor. Since the total number of small valued (i.e. $k = 1, 2, 3, 4$) coefficients are rougly equal between the two groups of α, we can infer that this lower bound is relatively close to the actual additional factor. Moreover, because we observe an equal number of small valued coefficients for both sets of α, the additional factor's value can range anywhere in the interval $[2^{816}, 4^{816}] \Rightarrow [4.37 \times 10^{245}, 1.91 \times 10^{491}]$. If this phenonom holds for a larger test group, then we possibly could believe the balancing hypothesis is valid.

Without loss of generality, we take $A = 1$ in Theorem 4.1.2; then for every α in the larger maximum set and for every α in the smaller maximum set, the empirically computed coefficient product is significantly less than e^{10^6}. Therefore, the six α, whose coefficient products were greater than Kuzmin's expected geometric mean, do behave wildly in their first 10^6 coefficients if this behavior is measured by the expectation given in Equation 4.4.30; but this wild behavior is not so extreme so as to characterize these six α as elements of the zero measure set for which Theorem 4.1.2 does not hold.

I will attempt to to provide some theoretical basis for the balancing hypothesis, which states the occurrence of an extremely large digit value will be balanced out by more than expected digits equal to 1 or 2. Recall we have from Kuzmin's Theorem, Theorem 4.4.2, and Equation 4.4.30:

$$\sqrt[n]{a_1 a_2 \cdots a_n} \to \prod_{r=1}^{\infty} \left\{ 1 + \frac{1}{r(r+2)} \right\}^{\frac{\log(r)}{\log 2}} = 2.68545 \qquad (4.4.37)$$

Let $K(l, n)$ be Kuzmin's expected number of coefficients $a_i = l$ in an expansion of length n. Suppose for some $\alpha = [0; a_1, \ldots, a_i, \ldots, a_n, \ldots] \in K$ and for α's continued fraction expansion to n coefficients, we have $a_i = k$ for some $i \leq n$, where k satisfies $n \ll \frac{1}{Prob(a_i=k)}$. It is clear there exists some $h \in N$, such that $(2.68545)^h \approx k$, and from Equation 4.4.37 and our assumption $\alpha \in K$, we know α's coefficient product must tend to $(2.68545)^n$. Therefore, if Equation 4.4.37 holds approximately for all n, then the length n expansion of α should possess roughly $K(l, n) + p$ coefficients with values $l = 1 \; or \; l = 2$, or a mix of both $l = 1, 2$, where $p \geq h$ (equality holds if all p 'balancing coefficients' have value 1). In other words, the coefficient $a_i = k$ is balanced out by a greater than expected number of coefficients with values 1 or 2 so that α's coefficient product tends to $(2.68545)^n$. But since Equation 4.4.37 holds in the limit and there is always some error associated with truncated continued fraction expansions, we may observe fewer balancing coefficients than p, and it is unlikely we will observe these balancing coefficients in the first n coefficients, but they should occur somewhere in the expansion of α. Finally it is important to mention the above argument does not preclude the frequent occurrence of large valued digits k_i close together, where the values k_i are distinct and close is determined relative to Kuzmin's expected frequency of each distinct k_i.

Chapter 5

Conclusion

This thesis has explored many facets of Kuzmin's Theorem. The first two chapters provided the necessary background to understand the theory in Chapters 3 and 4. All of the results presented in the early chapters were not absolutely essential, but every result or proof builds the reader's intuition about the behavior of continued fractions; this was the main goal of the first two chapters.

The most theoretical and detailed concept presented in this thesis was Kuzmin's Theorem, and while Kuzmin's Theorem is not elementary, I attempted to present it in full detail while keeping the arguments self-contained (within this thesis). The same Theorem was proved by Levy with a different error function. Although I neglected to include Levy's proof, I did provide a method for estimating both A and λ in his error function for a given set of α over a range of n. While A and λ are absolute constants in theory, as soon as we attempted to assign numerical values to these constants, we saw that both A and λ become functions of the range n, the beginning n, and the test set of α. Given that these absolute constants became functions of our data, I suspect we will encounter the same dilemna in answering what the minimum values of A and λ are such that Levy's inequality holds for a set of full measure.

Motivated by my work on continued fractions in 2003 [Mi], I wanted to find a more rigorous definition of the zero measure set Z, for which Kuzmin's Theorem does not hold. In addition to the many known subsets of Z, we used some theorems presented in Khintchine's work [Ki] to argue for the inclusion of two more sets in Z.

However, in finding these additional subsets, we realized we were on the road to finding Khintchine's constant, which allows us to subclassify irrational numbers. Finally, there may be additional theory to support the balancing hypothesis presented in Section 4.4.1, but more empirical work is needed to fully develop the intuition needed to prove or disprove this hypothesis.

Chapter 6

Appendix A

Bibliography

[De] R. DeVeaux, *The Probability Theory of Continued Fractions*, Senior Thesis. Princeton University, Princeton, NJ, 1973.

[Di] T. Dimofte, *Rational Shifts of Linearly Periodic Continued Fractions*, Junior Research Paper, Princeton University, February 12, 2003.

[Fi] D. Fishman, *Closed Form Continued Fraction Expansions of Special Quadratic Irrationals*, Princeton University, Princeton, May 19, 2003.

[Ki] A. Y. Khintchine, *Continued Fractions*, Third Edition, The University of Chicago Press, Chicago 1964.

[La] S. Lang, *Introduction to Diophantine Approximations*, Addison-Wesley, Reading, 1966.

[LT] S. Lang and H. Trotter, *Continued Fractions for some algebraic numbers*, J. Reine Angew. Math. **255**, 1972, 112-134.

[Le] P. Lévy, *Sur les lois de probabilité dont dependent les quotients complets et incomplets d'une fraction continue*, Bull. Soc. Math., **57**, 1929, $178 - 194$.

[Mi] M. Michelini *Continued Fraction Expansions of Algebraic Thrid Degree Primes*, Princeton University, Princeton, January 20, 2003.

[MT] S. Miller, R. Takloo-Bighash, H. Helfgott, and F. Spinu, *Junior Research Seminar: Diophantine Analysis and Approximations*, Princeton University, Princeton, 2002.

[Ni] I. Nivan, *Irrational Numbers*, Pub: The Mathematical Association of America, Dist: by John Wiley and Sons, Inc., Print: Quinn & Boden Company Inc., 1956.

[Ol] C.D. Olds, *Continued Fractions*, Pub: Random House. First printing by Yale University, 1963.