MAT 6932 - Number Theory Seminar

Jay Pantone
University of Florida

Last Edited: August 19, 2012

Contents

Chapter 1

Some Basic Theorems

1.1 Roots of Integers

Theorem 1: (Euclid) $\sqrt{2}$ is irrational.

Proof: Suppose toward a contradiction that $\sqrt{2}$ is rational. Then, $\sqrt{2} = \frac{p}{q}$ for some $p, q \in \mathbb{Z}$ with $\gcd(p, q) = 1$. Now,

$$\begin{aligned}
\sqrt{2} = \frac{p}{q} &\implies 2 = \frac{p^2}{q^2} \\
&\implies 2q^2 = p^2 \\
&\implies p^2 \text{ is even} \\
&\implies p \text{ is even} \\
&\implies p^2 = 4r^2 = 2q^2, \text{ for some } r \in \mathbb{Z} \\
&\implies 2r^2 = q^2 \\
&\implies q^2 \text{ is even} \\
&\implies q \text{ is even} \\
&\implies \gcd(p, q) \neq 1.
\end{aligned}$$

We have reached a contradiction. Thus $\sqrt{2}$ is not rational. $\square$.

Theorem 2: If $n \in \mathbb{Z}$ and n is not a perfect k^{th} power for some $k \in \mathbb{Z}^+$, then $\sqrt[k]{n}$ is irrational.

Proof: Suppose that $\sqrt[k]{n} = \frac{a}{b}$, for some $a, b \in \mathbb{Z}$. Define $\nu_p(m) :=$ the largest power of a prime p that divides m. So, $nb^k = a^k$. Now now that for all primes p, we have $\nu_p(a^k) \mid k$. Similarly, $\nu_p(b^k) \mid k$. Hence $\nu_p(n) \mid k$. Now, since <u>all</u> primes p divide n a multiple of k times, we have that n is a perfect k^{th} power. Thus we have proved the contrapositive. $\square$

Corollary: If $n, k \in \mathbb{Z}^+$, then either $\sqrt[k]{n} \in \mathbb{Z}$ or $\sqrt[k]{n}$ is irrational.

Theorem 3: (Rational Root Theorem) If $P(x) \in \mathbb{Z}[x]$ is written as $P(x) = a_0 + a_1x + \cdots + a_nx^n$, and if $P\left(\frac{p}{q}\right) = 0$, for $\frac{p}{q} \in \mathbb{Q}$ with $\gcd(p, q) = 1$, then $q \mid a_n$ and $p \mid a_0$.

Proof: Let $P(\frac{p}{q}) = a_0 + a_1\frac{p}{q} + \cdots + a_n\frac{p^n}{q^n}$. Then,

$$-a_0q^n = a_1q^{n-1}p + \cdots + a_np^n. \tag{3.1}$$

Now observe that the right-hand side of (3.1) is a multiple of p, and thus so is the left-hand side. Since $\gcd(p, q) = 1$, we have $p \mid a_0$. Similarly, since $q \mid$ LHS(3.1) and q divides all terms but a_np^n in the right-hand side of (3.1), we must have $q \mid a_np^n$. Since $\gcd(p, q) = 1$, it follows that $q \mid a_n$. □

Corollary: If $P(x) \in \mathbb{Z}[z]$ is <u>monic</u>, then the rational roots, if any, are integers which divide $a_0 = P(0)$.

Example: $P(x) = x^2 - 2$. By the **Rational Root Theorem**, there are no integer roots (we only need to test $\pm1, \pm2$). So, the two roots must be irrational, and thus $\sqrt{2}$ is irrational.

Exercise: Show that $\sqrt{2} + \sqrt{3}$ is irrational.

Archimedean Property: If α, β are positive reals, then there exists $n \in \mathbb{N}^+$ such that $n\alpha < \beta$.

Theorem 4: Between any two real numbers, there exists both a rational and an irrational number.

Proof: Without loss of generality, let $0 < \alpha < \beta$. Then, $\beta - \alpha > 0$. So, there exists $n \in \mathbb{Z}^+$ such that $n(\beta - \alpha) > 1$. Now, $n\beta > n\alpha + 1$. Since $n\beta$ and $n\alpha$ differ by more than 1, there is an integer m such that $n\alpha < m < n\beta$. Thus, $\alpha < \frac{m}{n} < \beta$.

To find an irrational number between α and β, find n such that $n(\beta-\alpha) > \sqrt{2}$ and proceed similarly. □

Corollary: Between any two rationals there exists an irrational. Between any two irrationals there exists a rational.

1.2 Some Theorems of Cantor

Theorem A: (Cantor) The set of rationals $\mathbb{Q}$ is countable.

Proof: First list the rationals in $[0,1]$ is lexicographical order by denominator then numerator:

$$\frac{0}{1}, \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \ldots.$$

To get all of $\mathbb{Q}$, take each fraction in the list, and follow it by its negative, its reciprocal, and its negative reciprocal:

$$\frac{0}{1}, \frac{1}{1}, -\frac{1}{1}, \frac{1}{2}, -\frac{1}{2}, \frac{2}{1}, -\frac{2}{1}, \frac{1}{3}, -\frac{1}{3}, \frac{3}{1}, -\frac{3}{1}, \ldots.$$

This is all of $\mathbb{Q}$, enumerated on the integers. Thus $\mathbb{Q}$ is countable. $\square$

Theorem B: (Cantor) $\mathbb{R}$ is uncountable.

Theorem C: (Cantor) The set $\mathbb{R} \setminus \mathbb{Q}$ of irrationals is uncountable.

Remark: Just because two things have the same cardinality doesn't mean it's easy to find a bijection between them. For example, consider the sets $[0,1]$ and $(0,1)$. To find a bijection between these two sets, pick a countably infinite set $X := \{0, 1, x_1, x_2, \ldots\} \subset [0,1]$. Then, let

$$\varphi(r) := \begin{cases} r, & r \notin X \\ x_1, & r = 0 \\ x_2, & r = 1 \\ x_{i+2}, & r = x_i \end{cases}.$$

Chapter 2

Decimal Representations

2.1 Decimal Representations

Definition: A decimal representation is a series

$$\sum_{i=1}^{\infty} \frac{a_i}{10^i},$$

where a_i satisfies $0 \leq a_i \leq 9$ and for infinitely many a_i, we have $a_i < 9$ (this prevents duplicate series such as $.6729999\ldots = .673$).

Remark: Note that:

$$0 \leq \sum_{i=1}^{\infty} \frac{a_i}{10^i} < \sum_{i=1}^{\infty} \frac{9}{10^i} = 1.$$

Theorem 5:

(i) Every decimal representation converges to a real number $\alpha \in [0, 1)$.

(ii) Every real $\alpha \in [0, 1)$ admits a decimal representation.

(iii) The decimal representation in (ii) is unique.

Proof: We've already shown **part (i)**. Next, denote $[x] :=$ the greatest integer $\leq x$ and $\{x\} := x - [x] \in [0, 1)$.

Uniqueness:

Let $\alpha = .a_1a_2a_3\ldots$. So, $10\alpha = a_1 + .a_2a_3\ldots$ and thus $a_1 = [10\alpha]$ and $\alpha_1 := .a_2a_3\ldots = \{10\alpha\}$. So, a_1 and α_1 are uniquely determined. Iterating, we see that all a_i are uniquely defined by

$$\begin{array}{l} a_{i+1} = [10\alpha_i] \\ \alpha_{i+1} = \{10\alpha_i\} \end{array} . \tag{5.1}$$

Existence

Given an α, define a_i as in (5.1) with $\alpha =: \alpha_0$. Clearly $0 \leq a_i \leq 9$, since $0 \leq \alpha_i < 1$. So $0 \leq 10\alpha_i < 10$. Now we claim $\alpha = .a_1a_2\ldots$, and we use induction. Note that $\alpha - .a_1 = \alpha - \frac{a_1}{10} = \frac{10\alpha - \alpha_1}{10}$.

Now we claim that $\alpha - .a_1a_2\ldots a_n = \dfrac{\{10a_{n-1}\}}{10^n}$. To see this, observe that

$$\alpha - .a_1a_2\ldots a_n = \alpha - \left(\frac{a_1}{10} + \cdots + \frac{a_{n-1}}{10^{n-1}}\right) - \frac{a_n}{10^n}.$$

By the Induction Hypothesis, this equals

$$\frac{\{10\alpha_{i-2}\}}{10^{n-1}} - \frac{a_n}{10^n} = \frac{10\alpha_{n-1} - [10\alpha_{n-1}]}{10^n} = \frac{\{10\alpha_{n-1}\}}{10^n}. \ \square$$

Now, by letting $n \to \infty$, we get existence.

Why do we still have that infinitely many $a_i < 9$? If $a_i = 9$ for all $i > N$, then α_i for $i > N$ is

$$\alpha_i = \frac{9}{10} + \frac{9}{10^2} + \cdots = 1.$$

But $\alpha_i = \{10\alpha_{i-1}\} < 1$, so this is impossible.

This completes the proof. $\square$

2.2 Decimal Expansions and a Criterion for Irrationality

Definition: A decimal expansion $.a_1a_2\cdots$ is periodic if there exist $m, n \in \mathbb{Z}$ such that $a_{t+n} = a_t$ for all $t > m$. The smallest such m is called the pre-period. The smallest such n is called the period.

We use the notation $.a_1a_2\cdots a_m\overline{a_{m+1}\cdots a_{m+n}}$.

Definition: A decimal expansion is purely periodic if $m = 0$.

Theorem 6: Let $\alpha \in \mathbb{R}$. Then, $\alpha \in \mathbb{Q}$ if and only if its decimal expansion is periodic.

Proof: ($\Longleftarrow$) First consider a purely periodic expansion $\beta = .\overline{c_1c_2\cdots c_n}$. Note that $10^n\beta = N + \beta$ for some $N \in \mathbb{N}$. Thus, $(10^n - 1)\beta \in \mathbb{N}$ and so $\beta = \frac{N}{10^n-1} \in \mathbb{Q}$. Next consider $\alpha := .a_1a_2\cdots a_m\overline{a_{m+1}\cdots a_n}$. Then, $10^m\alpha = N + \beta$ for some $N \in \mathbb{N}$ and purely period decimal expansion β. Since $\beta \in \mathbb{Q}$ by above, we have that $\alpha = \frac{N+\beta}{10^m} \in \mathbb{Q}$. $\square$

($\Longrightarrow$) If $\alpha = \frac{a}{b}$, with $\gcd(a, b) = 1$, consider the infinite sequence $\alpha, \{10\alpha\}, \{10^2\alpha\}, \ldots$. The values taken by the sequence are in the set $\{0, \frac{1}{b}, \frac{2}{b}, \ldots, \frac{b-1}{b}\}$. By the pigeon-hole principal, we get that $\{10^s\alpha\} = \{10^{s+r}\alpha\}$, for some $s, r \in \mathbb{N}$. Now,

$$\begin{aligned} \{10^s\alpha\} &= \{10^{s+r}\alpha\} \\ 10^s\alpha - [10^s\alpha] &= 10^{s+r}\alpha - [10^{s+r}\alpha] \\ [10^{s+r}\alpha] - [10^s\alpha] &= 10^s(10^r - 1)\alpha \end{aligned}$$

Let $N := [10^{s+r}\alpha] - [10^s\alpha] \in \mathbb{N}$. Let $N' \in \mathbb{Z}$ be such that $\dfrac{N}{10^r - 1} = N' + \dfrac{N_1}{10^r - 1}$, where $0 \le N_1 < 10^r - 1$.

Now note that $\dfrac{N_1}{10^r - 1} = \dfrac{N_1}{10^r} + \dfrac{N_1}{10^{2r}} + \dfrac{N_1}{10^{3r}} + \cdots$. Write N_1 in its decimal form, which is purely periodic. Then, $\alpha = (N' + \dfrac{N_1}{10^r - 1})\dfrac{1}{10^s}$ is periodic. $\square$

Corollary: α is irrational if and only if its decimal expansion is not periodic.

Example: In particular, $\alpha = .101001000100001\ldots$ is irrational.

Exercise: Let $\alpha = \frac{a}{b}$ be rational with $\gcd(a, b) = 1$. Write $b = 2^\epsilon \cdot 5^\nu \cdot M$, with $\gcd(M, 2) = \gcd(M, 5) = 1$.

Show that the pre-period of the decimal expansion of α equals $\max(\epsilon, \nu)$ and that the period of the decimal expansion of α equals the order of 10 in $\mathbb{Z}/M\mathbb{Z}$.

Cantor used decimal expansion to prove many things, such as a theorem we've already mentioned:

Theorem B: $\mathbb{R}$ is uncountable.

Proof: It suffices to show that $[0, 1) \in \mathbb{R}$ is uncountable. Suppose toward a contradiction that $[0, 1)$ is countable, then enumerate $[0, 1)$ as a sequence $\{\alpha_1, \alpha_2, \ldots\}$. Write the decimal expansion of each of these:

$$\begin{aligned} \alpha_1 &= .a_{11}a_{12}a_{13}\cdots \\ \alpha_2 &= .a_{21}a_{22}a_{23}\cdots \\ \alpha_3 &= .a_{31}a_{32}a_{33}\cdots \\ \vdots &= \quad \vdots \quad \ddots \end{aligned}$$

Consider $\beta \in [0,1)$ given by the decimal expansion $\beta = .b_1 b_2 \cdots$, where $b_i \neq a_{ii}$ and $b_i \in [0,8]$, for each i (this avoids a decimal expansion ending in all nines). The decimal expansion is unique from each element in the list above. Therefore, β is not in the enumerated list. This is a contradiction. So $\mathbb{R}$ is uncountable. $\square$

Theorem 7: Let $a_1, a_2, \ldots$ be an infinite sequence of $\mathbb{Z}$ with each $a_i \geq 2$. Then, for all $\alpha \in \mathbb{R}$ can be uniquely expressed as

$$\alpha = c_0 + \sum_{i=1}^{\infty} \frac{c_i}{a_1 \cdots a_i}, \text{ where } c_i \in \mathbb{Z}, \tag{7.1}$$

satisfying $0 \leq c_i \leq a_i - 1$, and infinitely many c_i satisfy $0 \leq c_i < a_i - 1$. Note that if all $a_i = 10$, then this is just the decimal expansion.

Proof: Exercise. Very similar to **Theorem 5**, with the following identity added:

$$\sum_{i=1}^{k} \frac{a_{n+i} - 1}{a_{n+1} \cdots a_{n+i}} = 1 - \frac{1}{a_{n+1} \cdots a_{n+k}}. \tag{7.2}$$

The connection between c_i and α is:

$$c_0 = [\alpha], \quad c_1 = \{\alpha\}, \quad c_i = [a_i \alpha_i], \quad \alpha_{i+1} = \{a_i \alpha_i\}.$$

Theorem 8: (Sufficient Condition for Irrationality) Let α, a_i, c_i be as in **Theorem 7**. Suppose that

(1) Infinitely many $c_i > 0$.

(2) For all primes p, it is true that p divides infinitely many a_i.

Then, α is irrational.

Proof: Suppose $\alpha = \frac{a}{b} \in \mathbb{Q}$, with $\gcd(a,b) = 1$. Then, for sufficiently large n, necessarily $b \mid a_1 \cdots a_n$ (by assumption (2)). Write

$$\alpha = c_0 + \sum_{i=1}^{\infty} \frac{c_i}{a_1 \cdots a_i} \tag{8.1}$$

and consider $\alpha - \left(c_0 + \sum_{i=1}^{n} \frac{c_i}{a_1 \cdots a_i} \right)$. Multiplying through by the product $a_1 \cdots a_n$, we have

$$(a_1 \cdots a_n) \left[\alpha - (c_0 + \sum_{i=1}^{n} \frac{c_i}{a_1 \cdots a_i}) \right] = \sum_{i=1}^{\infty} \frac{c_{n+i}}{a_{n+1} \cdots a_{n+i}}. \tag{8.2}$$

Suppose $\displaystyle\sum_{i=1}^{\infty} \frac{a_{n+i} - 1}{a_{n+1} \cdots a_{n+i}} = 1$, using (7.2) and letting $k \to \infty$.

Given that infinitely many $c_i < a_i - 1$, we have that $0 < \text{RHS}(8.2) < 1$. But, the left-hand side of (8.2) is an integer. This is a contradiction. Hence α is irrational. $\square$

2.3 Decimal Approximations

Fundamental Fact in the Theory of Irrationals:

Every nonzero integer is at least 1 in absolute value.

Remark: Given $\theta = \frac{a}{b} \in \mathbb{Q}$, consider a rational $\frac{p}{q} \neq \theta$. Use the convention that when writing a fraction, we let the denominator be positive. Then,

$$0 \neq \left|\theta - \frac{p}{q}\right| = \left|\frac{a}{b} - \frac{p}{q}\right| = \left|\frac{aq - pb}{bq}\right| = \frac{|aq - pb|}{bq} \geq \frac{1}{bq}.$$

The last step is due to the fact that $|aq - pb|$ is a positive integer greater than 0, and thus it's at least 1, i.e.,

$$\text{If } 0 \neq |q\theta - p|, \text{ then } |q\theta - p| \geq \frac{1}{b}. \tag{9.1}$$

Lemma 1: If θ is rational and $p, q \in \mathbb{Z}$, then either $q\theta - p = 0$ or $q\theta - p$ is bounded away from zero.

Corollary: If $\theta \in \mathbb{R}$ and there exist $p_n, q_n \in \mathbb{Z}$ such that $0 \neq q_n\theta - p_n$ and $q_n\theta - p_n \to 0$ as $n \to \infty$, then θ is irrational.

Remark: $\mathbb{Q}$ is dense in $\mathbb{R}$, i.e., given any $\theta \in \mathbb{R}$ there exist $\frac{p_n}{q_n} \in \mathbb{Q}$ such that $\frac{p_n}{q_n} \to \theta$ as $n \to \infty$. i.e., $\left|\theta - \frac{p_n}{q_n}\right| \to 0$.

If the stronger condition $|q_n\theta - p_n| \to 0$ holds, then θ is irrational (by contrapositive of **Lemma 1**).

Application: $\sqrt{2}$ is irrational.

Proof: (Alladi) Observe that $1 < \sqrt{2} < 2$. Thus, $0 < \sqrt{2} - 1 < 1$. Therefore, $0 \neq (\sqrt{2} - 1)^n \to 0$. Now, expand $(\sqrt{2} - 1)^n$ by the binomial theorem, and group terms so that $(\sqrt{2} - 1)^n = q_n\sqrt{2} - p_n$, for some $p_n, q_n \in \mathbb{Z}$.

Now, as $n \to \infty$, we have that $q_n\sqrt{2} - p_n \to 0$. Thus $\sqrt{2}$ is irrational by the above **Corollary**. □

Remark: Despite the above theorem, truncations of a decimal expansion are not usually strong enough to yield irrationality.

Application: (Euler) e is irrational.

Proof: We know that $e = \sum_{m=0}^{\infty} \frac{1}{m!}$. Consider $\frac{p_n}{q_n} = \sum_{m=0}^{n} \frac{1}{m!}$. Then, $0 < e - \frac{p_n}{q_n} = \sum_{m=n+1}^{\infty} \frac{1}{m!}$.

Next, note that $q_n = n!$, so

$$0 < |q_n e - p_n| = n! \cdot \sum_{m=n+1}^{\infty} \frac{1}{m!} = \frac{1}{(n+1)} + \frac{1}{(n+1)(n+2)} + \cdots < \frac{2}{n+1} \xrightarrow{n\to\infty} 0. \ \square$$

Remark: This proof works for the following reasons:

(1) The defining series for e is positive-termed, so each tail is non-zero.

(2) $q_n \mid q_{n+1}$.

Modifying the problem even a little bit ruins it. For example, Erdös asked:

$$\text{Is the number } \sum_{m=0}^{\infty} \frac{1}{m!+1} \text{ irrational?}$$

This is still unknown.

Remark: The corollary above gives us a sufficient condition. Is it also a necessary condition? Yes, and we prove this below.

Notation: We define $||x|| :=$ the distance from x to the nearest integer.

Theorem 10: (Dirichlet) Let θ be irrational, and let $Q > 1$ be an integer. Then, there exists $q \in \mathbb{Z}$ satisfying

$$0 < q \leq Q \text{ and } ||q\theta|| < \frac{1}{Q}.$$

Proof: Divide $[0,1)$ into Q subintervals: $[0, \frac{1}{Q})$, $[\frac{1}{Q}, \frac{2}{Q})$, $\ldots$, $[\frac{Q-1}{Q}, 1)$. Consider the following $Q+1$ numbers: $\{m\theta\}_{m=0,1,\ldots,Q}$. By the **Pigeon Hole Principle**, there exists an interval above such that two $\{m\theta\}$ are in it, i.e., there exist m, n with $m \neq n$ and $m < n$ such that

$$|\{m\theta\} - \{n\theta\}| < \frac{1}{Q}.$$

Note that $\{m\theta - n\theta\} = m\theta - [m\theta] - n\theta + [n\theta]$. Let $q := n - m$ and $p := [n\theta] - [m\theta]$. We then have:

$$0 \neq [q\theta - p] < \frac{1}{Q} \leq \frac{1}{2},$$

i.e., $\|q\theta\| < \frac{1}{Q}$. $\square$

Remark: We have proved the necessary condition, but we can find an even stronger result. Since θ is irrational, $\|q\theta\| \neq 0$. Thus as $q \to \infty$, there are infinitely many pairs (q, p) are generated. Consider this sequence (p_n, q_n).

Now, $0 < |q_n\theta - p_n| < \frac{1}{Q}$. So, $\left|\theta - \frac{p_n}{q_n}\right| < \frac{1}{q_n Q} < \frac{1}{q_n^2}$ (since $q \leq Q$). This yields the following theorems.

Theorem 11: $\theta \in \mathbb{R}$ is irrational if and only if there exist infinitely many (p_n, q_n) pairs such that $0 \neq |q_n\theta - p_n| \to 0$.

Theorem 12: If θ is irrational, then there exist $p_n, q_n \in Z$ for all $n \in \mathbb{N}$ such that

$$0 \neq \left|\theta - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2}. \tag{12.1}$$

Definition: The p_n, q_n satisfying (12.1) will be called a strong approximation.

Note: The tails (truncations) of the series for e, while good enough to prove irrationality, are *not* the strong approximations guaranteed to us by Dirichlet in **Theorem 12**.

Remark: On the other hand, if we write $(\sqrt{2}-1)^n$ as $q_n\sqrt{2}-p_n$, then these $q_n < (\sqrt{2}+1)^n$.

Proof: Exercise. Work with the binomial theorem, and calculate the q_n's.

Remark: Observe that $\sqrt{2}-1 = \frac{1}{\sqrt{2}+1}$. So, $q_n\sqrt{2}-p_n = (\sqrt{2}-1)^n = \frac{1}{(\sqrt{2}+1)^n} < \frac{1}{q_n}$ (with the last step by the above remark). Thus $\sqrt{2} - \frac{p_n}{q_n} < \frac{1}{q_n^2}$, and so these are our strong approximations.

2.4 Irrationality of π

Theorem 13: π is irrational.

Proof: First we state a lemma.

Lemma 2: Let $p(x)$ be a polynomial, and define

$$F(x) := p(x) - p''(x) + p^{(4)}(x) - p^{(6)}(x) + \cdots . \tag{13.1}$$

Then,

$$(F'(x)\sin(x) - F(x)\cos(x))' = p(x)\sin(x). \tag{13.2}$$

Proof: Exercise.

Now, suppose $\pi \in \mathbb{Q}$. Write π as $\frac{a}{b}$, with $\gcd(a,b) = 1$. Consider

$$I_n := \frac{1}{n!}\int_0^\pi \sin(x)x^n(a-bx)^n dx. \tag{13.3}$$

Let $P_n(x) := \frac{1}{n!}x^n(a-bx)^n$, so that $I_n = \int_0^\pi P_n(x)\sin(x)dx$. Now, by **Lemma 2**, we have that

$$I_n = \int_0^\pi P_n(x)\sin(x)dx = [F'(x)\sin(x) - F(x)\cos(x)]|_0^\pi = F(0) + F(\pi).$$

Well, $x^n(a-bx)^n \in \mathbb{Z}[x]$. Derivatives up to the $(n-1)^{\text{st}}$ will still have x factors, so $P_i(0) = 0$, and n^{th} derivatives and after will be divisible by $n!$ so that $P_i(0) \in \mathbb{Z}$. Thus we have that $F(0) \in \mathbb{Z}$.

Observe that

$$P(x) = \frac{b^n}{n!}x^n\left(\frac{a}{b} - x\right)^n = P\left(\frac{a}{b} - x\right).$$

Recall that $\pi = \frac{a}{b}$. So, $P(\pi) = P(\frac{a}{b} - \pi) = P(0)$. Thus by the above argument for $F(0)$, we have that $F(\pi) \in \mathbb{Z}$. Thus, $I_n = F(0) + F(\pi) \in \mathbb{Z}$.

Note that the integrand above is strictly positive, since $\sin(x), x, (a-bx)$ are all at least zero for the whole range and at some point strictly bigger than zero. Thus $I_n > 0$.

But also, $I_n < \frac{c^n}{n!} \to 0$ as $n \to \infty$, where $c = \max_{x\in\left[0,\frac{a}{b}\right]}(x(a-bx))$. So, $I_n < 1$ eventually as $n \to \infty$. This contradicts the above facts that $I_n \in \mathbb{Z}$ and $I_n > 0$. Therefore, π is irrational. $\square$

2.5 Irrationality of π^2

Theorem 14: π^2 is irrational.

Proof: (Beukers) First we prove a lemma.

Lemma 3: For integers $m \geq 0$, let $s_m := \int_0^1 t^m \sin(\pi t)dt$ and $c_m := \int_0^1 t^m \cos(\pi t)$. Then, there exist polynomials $P_m(x)$ and $Q_m(x)$ with integer coefficients such that

$$s_m = \frac{1}{\pi} P_m\left(\frac{1}{\pi^2}\right),$$

$$c_m = Q_m\left(\frac{1}{\pi^2}\right), \text{ and}$$

$$\deg P_m = \left[\frac{m}{2}\right], \ \deg Q_m = \left[\frac{m+1}{2}\right].$$

Proof: The starting values of are:

$$\left.\begin{aligned} s_0 = \frac{2}{\pi}, \ \ s_1 = \frac{1}{\pi} \\ c_0 = 0, \ \ c_1 = -\frac{2}{\pi^2} \end{aligned}\right\} \tag{14.1}$$

We have the reduction formulas:

$$s_m = \int_0^1 t^m \sin(\pi t)dt = \left.\frac{-t^m \cos(\pi t)}{\pi}\right|_0^1 + \frac{m}{\pi}\int_0^1 t^{m-1}\cos(\pi t)dt = \frac{1}{\pi} + \frac{m}{\pi}c_{m-1} \tag{14.2}$$

$$c_m = -\frac{m}{\pi}s_{m-1} \tag{14.3}$$

$$s_0 = \frac{2}{\pi}, \ \ s_1 = \frac{1}{\pi}, \ \ s_2 = \frac{1}{\pi} - \frac{4}{\pi^3}, \ \ s_3 = \frac{1}{\pi} - \frac{6}{\pi^3}$$

$$c_0 = 0, \ \ c_1 = -\frac{2}{\pi^2}, \ \ c_2 = -\frac{2}{\pi^2}, \ \ c_3 = -\frac{3}{\pi^2} + \frac{12}{\pi^4}$$

Complete the proof by induction on m.

Define the n^{th} Legendre polynomial by

$$P_n(x) = \frac{1}{n!} \cdot \frac{d^n}{dx^n}(x^n(1-x)^n). \tag{14.5}$$

Note that $P_n(x) \in \mathbb{Z}[x]$ and has degree n. Consider

$$I_n = \int_1^1 P_{2n}(t)\sin(\pi t)dt. \tag{14.6}$$

We use integration by parts by integrating the polynomial part and differentiating the trigonometric part. Note that at every stage in the integration by parts, there is x and $1-x$ as factors, so values at 0 and 1 vanish. Thus, after n integrations by parts, we arrive at the following:

$$I_n = \frac{(-1)^n \pi^{2n}}{(2n)!}\int_0^1 (\sin(\pi t))t^{2n}(1-t)^{2n}dt. \tag{14.7}$$

Note that in (14.7), the integrand is strictly positive. So I_n is nonzero. Also,

$$|I_n| < \left(\frac{\pi}{4}\right)^{2n} \cdot \frac{1}{(2n)!} \longrightarrow 0 \text{ (quite rapidly).} \tag{14.8}$$

By **Lemma 3**,

$$I_n = \frac{1}{\pi} Q_n^* \left(\frac{1}{\pi^2}\right). \tag{14.9}$$

If $\pi^2 = \frac{p}{q} \in \mathbb{Q}$, then since $I_n \neq 0$, we have $Q_n^* \left(\frac{1}{\pi^2}\right) \neq 0$. But then

$$|Q_n^* \left(\frac{1}{\pi^2}\right)| = |Q_n^* \left(\frac{q}{p}\right)| \geq \frac{1}{p^n} \text{ by the } \textbf{Fundamental Fact}.$$

So,

$$|I_n| \geq \frac{1}{\pi p^n}. \tag{14.10}$$

Now, notice that (14.8) and (14.10) are incompatible. Hence π^2 is irrational. □

Corollary: π is irrational.

Remark: Proving the irrationality of π^2 is stronger than proving the irrationality of π. We will see that to obtain irrationality measures for π, it is better to approach π directly and not through π^2.

Remark: The proof of **Theorem 13** (the irrationality of π) was following the method of Hermite in "Sur la fonction exponentielle" (1873).

2.6 Irrationality of the Trig Functions

Theorem 15: If $\theta \neq 0$ is rational, then $\cos\theta$ is irrational.

Proof: (Niven) Since $\cos\theta = \cos(-\theta)$, we concentrate on $\frac{\pi}{2} > \left[\theta = \frac{a}{b}\right] > 0$, with $\gcd(a,b) = 1$. Consider

$$P(x) := \frac{x^{p-1}(a-bx)^{2p}(2a-bx)^{p-1}}{(p-1)!}, \tag{15.1}$$

where p is a prime sufficiently large to be chosen later.

We rewrite $P(x)$ as

$$P(x) = \frac{(\theta-x)^{2p}\left(\theta^2-(\theta-x)^2\right)^{p-1}b^{3p-1}}{(p-1)!} \tag{15.2}$$

using completing the square. Note that

$$0 < P(x) < \frac{\theta^{2p}\theta^{2p-2}b^{3p-1}}{(p-1)!} \longrightarrow 0 \tag{15.3}$$

as $p \to \infty$, for $0 < x < \frac{a}{b} = \theta$.

As in **Theorem 13**, set

$$F(x) = P(x) - P''(x) + P^{(4)}(x) - P^{(6)}(x) + \cdots. \tag{15.4}$$

Therefore, by an earlier **Lemma**, $I := \int_0^\theta P(x)\sin(x)dx = \left[F'(x)\sin(x) - F(x)\cos(x)\right]\Big|_0^\theta$. Evaluating,

$$I = F'(\theta)\sin(\theta) - F(\theta)\cos(\theta) + F(0). \tag{15.5}$$

Notice that from (15.2) we see that P may be viewed as a function in $(\theta-x)^2$. Consequently, since

$$F'(x) = P'(x) - P'''(x) + P^{(5)}(x) - \cdots, \tag{15.6}$$

we have that

$$F'(\theta) = 0. \tag{15.7}$$

Thus $F'(\theta)\sin(\theta) = 0$.

Next, from the definition of P in (15.1), we have the following (15.8):

$$\left.\begin{array}{c} P^{(j)}(0) = 0, \text{ if } j \leq p-1. \\ P^{(j)}(0) \equiv 0 \bmod p, \text{ if } j > p-1. \\ P^{(p-1)}(0) = a^{2p}(2a)^{p-1} \end{array}\right\} \tag{15.8}$$

So, if $p > 1$, then by (15.8), we have $F(0) \not\equiv 0 \bmod p$, and $F(0) \in \mathbb{Z}$.

Let

$$q := F(0). \tag{15.9}$$

Finally, we note that

$$P(\theta-x) = \frac{x^{2p}\left(\theta^2-x^2\right)^{p-1}b^{3p-1}}{(p-1)!} = \frac{x^{2p}(a^2-b^2x^2)^{p-1}b^{p-1}}{(p-1)!}. \tag{15.10}$$

Thus, $P^{(j)}(\theta-x) \equiv 0 \bmod p$, for all p. Therefore, $F(\theta) \equiv 0 \bmod p$. Let m be such that

$$F(\theta) = mp. \tag{15.11}$$

Thus,

$$I = -mp\cos(\theta) + q. \tag{15.12}$$

Remember that $q \neq 0$ is an integer. So $|q| \geq 1$. (If $-mp\cos(\theta) = 0$, then we have a contradiction since $I \to 0$ and $q \in \mathbb{Z}$. Thus $-mp\cos(\theta) \neq 0$.)

Recall that $I \to 0$ as $p \to \infty$ and $I \neq 0$ (since $I > 0$ because it has positive integrand due to the range restriction of θ).

Thus $-mp\cos(\theta) \neq 0$. Consequently, if $\cos(\theta)$ is rational, then we have an integral linear combination of $\cos(\theta)$ and 1 that is non-zero and goes to zero. If $\cos(\theta)$ is rational, this violates **Dirichlet's Criterion**. Thus $\cos(\theta)$ is irrational. □

Corollary 6: If $\theta \neq 0$ is rational, then $\sin\theta$ is irrational.

Proof: We use the identity $\cos 2\theta = 1 - 2\sin^2\theta$ to deduce that if $\sin\theta$ is rational, then so is $\cos 2\theta$, but $\cos 2\theta$ is not rational if $\theta \neq 0$. □

Corollary 7: If $\theta \neq 0$ is rational, then $\tan\theta$ is irrational.

We use the identity $\cos 2\theta = \dfrac{1 - \tan^2\theta}{1 + \tan^2\theta}$ to deduce that if $\tan\theta$ is rational, then so is $\cos 2\theta$, but $\cos 2\theta$ is not rational if $\theta \neq 0$. □

Corollary 8: As above, $\sec\theta$, $\csc\theta$, and $\cot\theta$ are all irrational if $\theta \neq 0$ and $\theta \in \mathbb{Q}$.

Remark: No one currently knows how to deduce the irrationality of all the trig functions at non-zero rational arguments by starting with the irrationality of $\sin\theta$ or $\tan\theta$ for $\theta \in \mathbb{Q}$, $\theta \neq 0$.

Remark: The identity $\cos 2\theta = 1 - 2\sin^2\theta$, along with **Theorem 15** actually yields the irrationality of $\sin^2\theta$ for non-zero rational θ. Likewise for $\tan^2\theta$. This in turn yields the irrationality of $\cos^2\theta$ for non-zero rational θ by the identity $\sin^2\theta + \cos^2\theta = 1$.

Remark: Our method of proof of **Theorem 15** was an extension / strengthening of the method used to prove the irrationality of π. Indeed, the irrationality of π follows from **Theorem 15**, because $\cos(\pi) = -1$ is rational, thus π is not rational.

2.7 Irrationality of the Hyperbolic Trig Functions

Theorem 16: If $\theta \in \mathbb{Q}$ is non-zero, then $\cosh\theta$ is irrational.

Proof: Since $\cosh\theta$ is an even function, we focus on $\theta > 0$. Let $\theta = \frac{a}{b}$, for some $a, b \in \mathbb{Z}$, with $\gcd(a,b) = 1$. As earlier, let $P(x) := \dfrac{x^{p-1}(a-bx)^{2p}(2a-bx)^{p-1}}{(p-1)!}$, where p is a large prime to be chosen later. Now, define $F(x) := P(x) + P''(x) + P^{(4)}(x) + P^{(6)} + \cdots$. Observe that

$$\begin{aligned}[F(x)\cosh(x) - F'(x)\sinh(x)]' &= F'(x)\cosh(x) + F(x)\sinh(x) - F''(x)\sinh(x) - F'(x)\cosh(x) \\ &= (F(x) - F''(x))\sinh(x). \end{aligned} \tag{16.1}$$

and we see through telescoping that

$$(F(x) - F''(x))\sinh(x) = P(x)\sinh(x). \tag{16.2}$$

Therefore,

$$I := \int_0^\theta P(x)\sinh(x)dx = F(x)\cosh(x) - F'(x)\sinh(x)\big|_0^\theta = F(\theta)\cosh\theta - F'(\theta)\sinh\theta - F(0). \tag{16.3}$$

As before, $P(x) > 0$ for $0 < x < \frac{a}{b} = \theta$. Also, $\sinh x > 0$ for $x > 0$. Hence the integrand in (16.3) is strictly positive. Thus $I > 0$. As before, $F'(\theta) = 0$, and $F(0)$ and $F(\theta)$ are integers. (See **Theorem 14**.)

Suppose that $\cosh\theta$ is rational. Let $\cosh\theta = \cosh\frac{a}{b} = \frac{\ell}{k}$, for $\ell, k \in \mathbb{Z}$. Now,

$$I = \frac{\ell}{k}F(\theta) - F(0). \tag{16.4}$$

We know by earlier arguments that $F(0) \neq 0$ if we choose p to be large enough. Note that $0 < I < \theta \cdot \dfrac{e^\theta - e^{-\theta}}{2} \cdot \dfrac{\theta^{4p-2}b^{3p-1}}{(p-1)!} \longrightarrow 0$ (That value is the max value of the integrand (max value of $\sinh x$ times the max value of $P(x)$) multiplied by the width of integration.)

So, from (16.4), we see that $\frac{\ell}{l}F(\theta) \neq 0$. Rewrite (16.4) as $kI = \ell F(\theta) - kF(0)$. Thus $kI \in \mathbb{Z}$. But, by choosing p large enough, we can make $kI < 1$. Since $kI > 0$, this is a contradiction. Hence $\cosh\theta = \cosh\frac{a}{b}$ is irrational. □

Corollary 9: If θ is a nonzero rational, then $\sinh\theta$ is irrational.

Proof: The identity being used is $1 + 2\sinh^2\theta = \cosh 2\theta$. □

Corollary 10: Similarly, $\tanh\theta$ is irrational.

Theorem 17: If θ is a nonzero rational, then e^θ is irrational.

Theorem 18: If $\theta \neq 1$ is rational, then $\log(\theta)$ is irrational.

Proof: $\exp : x \mapsto e^x$ is a function that sends $\mathbb{Q} \setminus \{0\}$ to a subset of $\mathbb{R} \setminus \mathbb{Q}$, so the inverse function log maps into $\mathbb{Q}$ only from 1 and some of $\mathbb{R} \setminus \mathbb{Q}$. Thus log maps from $\mathbb{Q} \setminus \{1\}$ into $\mathbb{R} \setminus \mathbb{Q}$. □

Chapter 3

Engel Series Expansions

3.1 Construction

Start with $x_0 \in (0, 1]$. Define

$$u_0 := \left[\frac{1}{x_0}\right] + 1. \tag{19.1}$$

Note that

$$\frac{1}{x_0} \geq 1,$$

so thus

$$\left[\frac{1}{x_0}\right] \geq 1.$$

Therefore, $u_0 \geq 2$, and of course $u_0 \in \mathbb{Z}$. Also, $\dfrac{1}{x_0} < \left[\dfrac{1}{x_0}\right] + 1 = u_0$. So, $u_0 x_0 > 1$. Now we define

$$x_1 := u_0 x_0 - 1. \tag{19.2}$$

Observe from (19.1) that

$$\frac{1}{x_0} < u_0 < \frac{1}{x_0} + 1.$$

Therefore,

$$0 < u_0 x_0 - 1 = x_1 \leq x_0 \leq 1. \tag{19.3}$$

Thus $x_1 \in [0, 1)$ like x_0, and $x_1 \leq x_0$. By iterating, we will get an infinite sequence

$$0 < \ldots \leq x_2 \leq x_1 \leq x_0 \leq 1. \tag{19.4}$$

Obviously, $\lim\limits_{n\to\infty} =: \ell$ exists, and $\ell \in [0, 1]$. The question is: when does $\ell = 0$ and when is $\ell > 0$? The answer to this question provides a criterion for irrationality.

Iteration Formula:

$$u_n = \left[\frac{1}{x_n}\right] + 1, \qquad x_{n+1} = u_n x_n - 1. \tag{19.6}$$

From (19.4), it follows that $\frac{1}{x^n} \leq \frac{1}{x_{n+1}}$, and therefore:

$$2 \leq u_0 \leq u_1 \leq u_2 \leq \ldots \text{ is a sequence of integers.} \tag{19.7}$$

Remark: Note from (19.6) that

$$x_n \to 0 \text{ if and only if } u_n \to \infty. \tag{19.8}$$

Equivalently,

$$x_n \to \ell \text{ for } \ell > 0 \text{ if and only if } u_n \to L \in \mathbb{Z}^+, \text{ in which case } u_n = L, \text{ for all } n \geq N. \tag{19.9}$$

Remark: We see from (19.6) that

$$\begin{aligned} x_0 &= \frac{1 + x_1}{u_0} \\ &= \frac{1}{u_0} + \frac{x_1}{u_0} \\ &= \frac{1}{u_0} + \frac{1 + x_2}{u_0 u_1} \\ &= \frac{1}{u_0} + \frac{1}{u_0 u_1} + \frac{x_2}{u_0 u_1}. \end{aligned}$$

Iterating this, we get that

$$x_0 = \frac{1}{u_0} + \frac{1}{u_0 u_1} + \frac{1}{u_0 u_1 u_2} + \cdots + \frac{1}{u_0 u_1 u_2 \cdots u_n} + \frac{x_{n+1}}{u_0 u_1 u_2 \cdots u_n}. \tag{19.10}$$

Letting $n \to \infty$ in (19.10), we arrive at the next theorem.

3.2 Some Theorems

Theorem 19-1: Let $x_0 \in (0,1]$ and define u_n by (19.6). Then:

$$x_0 = \frac{1}{u_0} + \frac{1}{u_0 u_1} + \frac{1}{u_0 u_1 u_2} + \cdots . \tag{19.11}$$

This is called the Engel Series for x_0.

Theorem 19-2: The representation of a real number $x_0 \in (0,1]$ by an Engel series is unique. That is, if we write x_0 as in (19.11) and if the u_n satisfy (19.7), then u_n are given by (19.6).

Proof: We have that the u_n are increasing, and $u_0 \geq 2$. From (19.11) we get that

$$0 < x_1 := u_0 x_0 - 1 = \frac{1}{u_1} + \frac{1}{u_1 u_2} + \cdots \leq \frac{1}{u_0} + \frac{1}{u_0 u_1} + \cdots \leq x_0 \leq 1.$$

Therefore:

$$\frac{1}{x_0} < u_0 \leq \frac{1}{x_0} + 1.$$

But, $u_0 \in \mathbb{Z}$. Thus,

$$u_0 = \left[\frac{1}{x_0}\right] + 1.$$

Hence, u_0 is uniquely determined.

Iteratively, it can be shown that u_n is given by (19.6) and thus is uniquely determined. □

Remark: In the Cantor Representation (i.e., the decimal representation but with changing base at each decimal place - more on this shortly), we set:

$$x_0 = \sum \frac{a_m}{c_1 c_2 \cdots c_m},$$

with the sequence $\{c_i\}_{i \in \mathbb{N}}$ specified ahead of time. Then, each x_0 determines a unique sequence of "digits" $\{a_m\}$.

In the Engel Series Representation, the a_m are specified to be all 1. Then, we have that each x_0 determines a unique sequence $\{c_i\}_{i \in \mathbb{N}}$.

Theorem 20: Let $x_0 \in (0,1]$. Then $x_0 \in \mathbb{Q}$ if and only if in the Engel Series of x_0, the sequence $\{u_n\}$ is eventually constant.

Proof: ($\Longleftarrow$) Let $u_n = L$, for all $n \geq N$. Then,

$$\begin{aligned} x_0 &= \frac{1}{u_0} + \frac{1}{u_0 u_1} + \cdots + \frac{1}{u_0 u_1 \cdots u_{N-1}} + \frac{1}{u_0 u_1 \cdots u_{N-1}} \left(\frac{1}{L} + \frac{1}{L^2} + \frac{1}{L^3} + \cdots \right) \\ &= \frac{1}{u_0} + \frac{1}{u_0 u_1} + \cdots + \frac{1}{u_0 u_1 \cdots u_{N-1}} + \frac{1}{u_0 u_1 \cdots u_{N-1}} \left(1 + \frac{1}{L-1} \right), \end{aligned}$$

which is clearly rational.

($\Longrightarrow$) (version 1) Suppose $x_0 \in \mathbb{Q} \cap (0,1]$. Write $x_0 = \frac{A_0}{B_0}$, for $A_0, B_0 \in \mathbb{Z}^+$. Note that

$$0 < A_0 \leq B_0. \tag{20.1}$$

Use the division algorithm:

$$B_0 = A_0 Q_0 + R_0, \text{ with } 0 \leq R_0 < A_0.$$

Now,

$$\begin{aligned} x_1 &= u_0 x_0 - 1 \\ &= \left(\left[\frac{1}{x_0} \right] + 1 \right) x_0 - 1 \\ &= \left(\left[\frac{B_0}{A_0} \right] + 1 \right) \frac{A_0}{B_0} - 1 \\ &= (Q_0 + 1) \frac{A_0}{B_0} - 1 \\ &= \frac{Q_0 A_0 + A_0 - B_0}{B_0} \\ &= \frac{A_0 - R_0}{B_0} \\ &= \frac{A_1}{B_0}, \end{aligned} \tag{20.2}$$

for some $0 < A_1 \leq A_0$, with $A_1 \in \mathbb{Z}^+$.

Now, by iteration, we have

$$x_n = \frac{A_n}{B_0}, \tag{20.3}$$

with $0 < \cdots \leq A_2 \leq A_1 \leq A_0$, with the A_n decreasing, but staying positive integers.

Thus $A_n = D$, for all $n \geq N$, and some $D \in \mathbb{Z}^+$. This implies that $\ell > 0$, and so $u_n = L$, for all $n \geq N$.

($\Longrightarrow$) (version 2) The contrapositive is the statement that if $u_n \to \infty$, then x_0 is irrational. From (19.10), we get that

$$0 < x_0 - \sum_{m=0}^{n} \frac{1}{u_0 u_1 \cdots u_m} = \frac{x_{n+1}}{u_0 u_1 \cdots u_{n+1}}. \tag{20.7}$$

Now, $u_n \to \infty \iff x_{n+1} \to 0$, by (19.6).

Rewrite (20.7) as

$$0 < q_n x_0 - p_n = x_{n+1}, \tag{20.8}$$

where

$$\begin{aligned} q_n &= u_0 \cdots u_n, \\ p_n &= u_0 \cdots u_n \sum_{m=0}^{n} \frac{1}{u_0 \cdots u_m}, \end{aligned}$$

with $q_n, p_n \in \mathbb{Z}$.

In (20.8), we have $x_{n+1} \to 0$, therefore by the Dirichlet Criterion, x_0 is irrational. □

Application: $\cosh(\sqrt{2})$ is irrational.

Proof: First, recall that

$$\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots .$$

Therefore,

$$\begin{aligned} \cosh(\sqrt{2}) &= \sum_{n=0}^{\infty} \frac{2^n}{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1) \cdot 2 \cdot 4 \cdot \ldots \cdot 2n} \\ &= \sum_{n=0}^{\infty} \frac{1}{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1) \cdot 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n} \end{aligned}$$

i.e., $\cosh(\sqrt{2}) - 2 = \sum_{m=1}^{\infty} \frac{1}{u_0 \cdots u_m}$ is an Engel Series, with $u_m = (m+2)(2m+3)$, for $m \geq 0$. Hence $\cosh(\sqrt{2}) - 2$ is irrational and so is $\cosh(\sqrt{2})$. $\square$

Corollary 14: $e^{\sqrt{2}}$ is irrational.

Proof: $\cosh(\sqrt{2}) = \frac{e^{\sqrt{2}} + \frac{1}{e^{\sqrt{2}}}}{2}$. $\square$

Exercise: Let $F_n = 2^{2^n} + 1$ be the n^{th} Fermat number. Then

$$\sum \frac{1}{F_n} \text{ is irrational.} \qquad (\bigstar)$$

Hint: First prove that $\sum_{n=0}^{\infty} \frac{1}{2^{2^n} - 1}$ is irrational using Engel Series, then connect this back to ($\bigstar$).

Chapter 4

Cantor Product Representation

4.1 Construction

Construction: Given $2 > \alpha_0 > 1$, define

$$\left.\begin{aligned} q_0 &= \left[\frac{\alpha_0}{\alpha_0 - 1}\right] \in \mathbb{Z} \\ \alpha_1 &= \frac{\alpha_0}{1 + \frac{1}{q_0}} \end{aligned}\right\} \tag{21.1}$$

Therefore,

$$q_0 \leq \frac{a_0}{a_0 - 1} < q_0 + 1. \tag{21.2}$$

From the second inequality in (21.2), it follows that

$$1 + \frac{1}{\alpha_0 - 1} < q_0 + 1 \iff \frac{1}{\alpha_0 - 1} < q_0 \iff \frac{1}{q_0} < \alpha_0 - 1 \iff 1 + \frac{1}{q_0} < \alpha_0. \tag{21.3}$$

From (21.1) and (21.3), we deduce that

$$1 < \alpha_1 < \alpha_0.$$

Now, we can iteratively define:

$$q_n = \left[\frac{\alpha_n}{\alpha_n - 1}\right], \qquad \alpha_{n+1} = \frac{\alpha_n}{1 + \frac{1}{q_n}}, \tag{21.4}$$

to get a sequence

$$\alpha_0 > \alpha_1 > \alpha_2 > \cdots > 1. \tag{21.5}$$

As with Engel Series, we have that

$$\lim_{n \to \infty} \alpha_n = \alpha \geq 1 \text{ exists.}$$

The natural question is when does this limit equal 1, and when is it strictly greater than 1?

Since the function $\frac{x}{x+1}$ is decreasing for $x > 1$, we have that

$$\frac{\alpha_0}{\alpha_0 - 1} < \frac{\alpha_1}{\alpha_1 - 1} < \frac{\alpha_2}{\alpha_2 - 1} < \cdots. \tag{21.6}$$

Therefore,

$$q_0 \leq q_1 \leq q_2 \leq \cdots. \tag{21.7}$$

Similarly to Engel series, if the q series stabilizes, then the above limit α is strictly greater than 1. From (21.1) and the first inequality in (21.2), we get the following:

$$\alpha_1 \leq \frac{\alpha_0}{1+\dfrac{\alpha_0-1}{\alpha_0}} = \frac{\alpha_0^2}{2\alpha_0-1}. \tag{21.8}$$

Therefore,

$$\frac{\alpha_1}{\alpha_1-1} \geq \frac{\alpha_0^2}{2\alpha_0-1}\cdot\frac{1}{\frac{\alpha_0^2}{2\alpha_0-1}-1} = \frac{\alpha_0^2(2\alpha_0-1)}{(2\alpha_0-1)(\alpha_0^2-2\alpha_0-1)} = \left(\frac{\alpha_0}{\alpha_0-1}\right)^2. \tag{21.9}$$

Hence,

$$q_1 = \left[\frac{\alpha_1}{\alpha_1-1}\right] \geq \left[\frac{\alpha_0^2}{(\alpha_0-1)^2}\right] \geq \left[\frac{\alpha_0}{\alpha_0-1}\right]^2 = q_0^2. \tag{21.10}$$

Thus, the iteration yields

$$\frac{\alpha_{n+1}}{\alpha_{n+1}-1} \geq \frac{\alpha_n^2}{(\alpha_n-1)^2}, \tag{21.11}$$

and

$$q_{n+1} \geq q_n^2. \tag{21.12}$$

Rewrite (21.1) as $\alpha_0 = \left(1+\frac{1}{q_0}\right)\alpha_1 = \left(1+\frac{1}{q_0}\right)\left(1+\frac{1}{q_1}\right)\alpha_2 = \ldots$. So, we can write

$$\alpha_0 = \prod_{j=0}^{n}\left(1+\frac{1}{q_j}\right)\cdot a_{n+1}. \tag{21.13}$$

So, since $q_0 > 1$ by definition, we have that $q_n \to \infty$, and so

$$\alpha_{n+1} \to 1. \tag{21.14}$$

Therefore, the case where $\lim\limits_{n\to\infty}\alpha_n > 1$ is not possible.

4.2 Some Theorems

Theorem 21-1: Every real number can be written as a product

$$\alpha_0 = \prod_{n=0}^{\infty} \left(1 + \frac{1}{q_n}\right), \tag{21.15}$$

where the q_n are positive integers satisfying $q_{n+1} \geq q_n^2$, and $q_0 > 1$.

Theorem 21-2: If a real number $\alpha_0 > 1$ is written as $\alpha_0 = \prod_{n=0}^{\infty} \left(1 + \frac{1}{q_n}\right)$, with $q_{n+1} \geq q_n^2$, and $q_0 > 1$, then the q_n are uniquely determined.

Proof: Write

$$\alpha_0 = \left(1 + \frac{1}{q_0}\right) \alpha_1, \tag{21.16}$$

with

$$\alpha_1 = \prod_{n=1}^{\infty} \left(1 + \frac{1}{q_n}\right). \tag{21.17}$$

Thus, $1 < \alpha_1 < \alpha_0$. Also,

$$\alpha_0 \leq \left(1 + \frac{1}{q_0}\right)\left(1 + \frac{1}{q_0^2}\right)\left(1 + \frac{1}{q_0^4}\right)\left(1 + \frac{1}{q_0^8}\right)\cdots.$$

Also,

$$\left(1 + \frac{1}{q_0}\right)\left(1 + \frac{1}{q_0^2}\right)\left(1 + \frac{1}{q_0^4}\right)\left(1 + \frac{1}{q_0^8}\right)\cdots = \frac{1}{1 - \frac{1}{q_0}} = \frac{q_0}{q_0 - 1}. \tag{21.18}$$

The equality above is due to the analytic identity:

$$\prod_{k=0}^{\infty} \left(1 + x^{2^k}\right) = \frac{1}{1 - x}.$$

Now, from (21.18) it follows that

$$\begin{aligned} \alpha_0 \leq 1 + \frac{1}{q_0 - 1} &\iff \alpha_0 - 1 \leq \frac{1}{q_0 - 1} \\ &\iff q_0 - 1 \leq \frac{1}{\alpha_0 - 1} \\ &\iff q_0 \leq 1 + \frac{1}{\alpha_0 - 1} = \frac{\alpha_0}{\alpha_0 - 1}. \end{aligned} \tag{21.19}$$

Also from (21.5), we have that

$$\alpha_0 > 1 + \frac{1}{q_0} \Longrightarrow \frac{\alpha_0}{\alpha_0 - 1} < \frac{1 + \frac{1}{q_0}}{\frac{1}{q_0}}. \tag{21.20}$$

So, from (21.19) and (21.20), we get $q_0 = \left[\frac{\alpha_0}{\alpha_0 - 1}\right]$. Thus q_0 is uniquely defined, and by iteration, all q_i are uniquely defined. □

Theorem 22: If $2 > \alpha > 1$ is represented as $\alpha = \prod_{n=0}^{\infty}\left(1+\frac{1}{q_n}\right)$ with $q_{n+1} \geq q_n^2$ and $q_n \in \mathbb{Z}$ and $q_0 > 1$, then $\alpha \in \mathbb{Q}$ if and only if $q_{n+1} = q_n^2$ for n sufficiently large.

Proof: ($\Longleftarrow$) Suppose $q_{n+1} = q_n^2$ for all $n \geq N$. Then,

$$\begin{aligned}\alpha_0 &= \left(1+\frac{1}{q_0}\right)\left(1+\frac{1}{q_1}\right)\left(1+\frac{1}{q_2}\right)\cdots\left(1+\frac{1}{q_{N-1}}\right)\cdot\prod_{k=0}^{\infty}\left(1+\frac{1}{q_N^{2^k}}\right)\\ &= \left(1+\frac{1}{q_0}\right)\left(1+\frac{1}{q_1}\right)\left(1+\frac{1}{q_2}\right)\cdots\left(1+\frac{1}{q_{N-1}}\right)\cdot\frac{1}{1-\frac{1}{q_N}} \in \mathbb{Q}. \end{aligned} \tag{22.1}$$

($\Longrightarrow$) Suppose $\alpha_0 \in \mathbb{Q}$ and α_0 given by the product (21.15) with q_n satisfying (21.12). Then $\alpha_n \in \mathbb{Q}$, with α_n defined by:

$$\alpha_n = \prod_{m=n}^{\infty}\left(1+\frac{1}{q_m}\right). \tag{22.2}$$

Write

$$\alpha_n = \frac{a_n}{b_n}, \tag{22.3}$$

with $a_n, b_n, \in \mathbb{Z}$ and $\gcd(a_n, b_n) = 1$. Since

$$\alpha_{n+1}\cdot\left(1+\frac{1}{q_n}\right) = a_n, \tag{22.4}$$

we have that

$$\frac{a_n}{b_n} = \frac{a_{n+1}}{b_{n+1}}\left(1+\frac{1}{q_n}\right) = \frac{a_{n+1}(q_n+1)}{b_{n+1}q_n} \iff \frac{a_{n+1}}{b_{n+1}} = \frac{a_n q_n}{b_n(q_n+1)}. \tag{22.5}$$

Therefore,

$$\begin{matrix} a_n q_n = a_{n+1}d_n \\ b_n(q_n+1) = b_{n+1}d_n \end{matrix}\ ,\ d_n \in \mathbb{Z}^+ \tag{22.6}$$

Now by subtraction:

$$(a_n - b_n)q_n - b_n = d_n(a_{n+1} - b_{n+1}) \geq a_{n+1} - b_{n+1}. \tag{22.7}$$

From our recursive formula, we get:

$$q_n \leq \frac{\alpha_n}{\alpha_n - 1} = \frac{a_n}{b_n}\cdot\frac{1}{(a_n-1)} = \frac{a_n}{a_n - b_n} \iff a_n(a_n - b_n) \leq a_n. \tag{22.8}$$

Thus by (22.7) and (22.8), we have:

$$a_n - b_n \geq a_{n+1} - b_{n+1} > 0. \tag{22.9}$$

Now, $\{a_n - b_n\}$ is a decreasing sequence of positive integers, so it must stablize to some $k = a_n - b_n$ for all $n \geq N$. Therefore, $q_n k - b_n = d_n k$ for $n \geq N$. Hence $b_n \equiv 0 \bmod k$, for all $n \geq N$. But $k = a_n - b_n$, so $a_0 \equiv 0 \bmod k$ for all $n \geq N$.

Since $\gcd(a_n, b_n) = 1$, we have that $k = 1$. Thus $a_n - b_n = 1$ for all $n \geq N$. Therefore, $\alpha_n = \frac{a_n}{b_n} = \frac{a_n}{a_n - 1}$, which implies that $\mathbb{Z} \ni a_n = \frac{\alpha_n}{\alpha_n - 1}$. Thus $q_n = \left[\frac{\alpha_n}{\alpha_n - 1}\right] = a_n$. Also, $q_n a_n = a_{n+1}d_n$, and therefore $q_n^2 = q_{n+1}d_n \geq q_{n+1}$. But, $q_{n+1} \geq q_n^2$. Therefore, $q_{n+1} = q_n^2$, for all $n \geq N$. $\square$

Remark: The theorem we stated is for real numbers $1 < r < 2$. If $\alpha_0 > 2$, then $q_0 = 1$, and thus $\alpha_1 = \frac{\alpha_0}{2}$. Eventually, by the same iteration, we get an α_N such that $1 < \alpha_N < 2$, and we just start here.

4.3 Evaluation of a Special Cantor Product

Theorem: Let $q_0 > 1$ and q_n be defined by

$$q_{n+1} = 2q_n^2 - 1, \tag{23.1}$$

for all $n \geq 0$. Then,

$$\alpha_0 = \prod_{n=0}^{\infty} \left(1 + \frac{1}{q_n}\right) = \sqrt{\frac{q_0 + 1}{q_0 - 1}}. \tag{23.2}$$

Proof: By (23.1),

$$\begin{aligned}
\left(1 + \frac{1}{q_n}\right)\left(1 + \frac{1}{q_{n+1}}\right) &= \frac{q_n + 1}{q_n} \cdot \frac{q_{n+1} + 1}{q_{n+1}} \\
&= \frac{(q_n + 1)2q_n^2}{q_n \cdot q_{n+1}} \\
&= \frac{q_n - 1}{q_n - 1} \cdot \frac{(q_n + 1)2q_n^2}{q_n \cdot q_{n+1}} \\
&= \frac{(q_n - 1)^2 \cdot 2q_n}{q_{n+1}(q_n - 1)} \\
&= \frac{(q_{n+1})q_n}{q_{n+1}(q_n - 1)} \\
&= \frac{1 - \frac{1}{q_{n+1}}}{1 - \frac{1}{q_n}}.
\end{aligned} \tag{23.3}$$

Taking the product of both values:

$$\prod_{n=0}^{\infty} \left(1 + \frac{1}{q_n}\right) \prod_{n=0}^{\infty} \left(1 + \frac{1}{q_{n+1}}\right) = \prod_{n=0}^{\infty} \frac{1 - \frac{1}{q_{n+1}}}{1 - \frac{1}{q_n}}. \tag{23.4}$$

Thus,

$$\frac{\alpha_0^2}{1 + \frac{1}{q_0}} = \frac{1}{1 - \frac{1}{q_0}} \Longrightarrow \alpha_0^2 = \frac{q_0 + 1}{q_0 - 1},$$

and this yields (23.2). □

4.4 Connection With Engel Series

Lemma: If $q_{n+1} = 2q_n^2 - 1$, then

$$q_n - \sqrt{q_n^2 - 1} = \frac{1}{2q_n} + \frac{1}{2q_n}\left(q_{n+1} - \sqrt{q_{n+1}^2 - 1}\right).$$

Proof: Note that $q_{n+1}^2 - 1 = (2q_n^2 - 1)^2 - 1 = 4q_n^4 - 4q_n^2$. Thus,

$$\begin{aligned}
\frac{1}{2q_n} + \frac{1}{2q_n}\left(q_{n+1} - \sqrt{q_{n+1}^2 - 1}\right) &= \frac{1}{2q_n} + \frac{1}{2q_n}\left(2q_n^2 - 1 - \sqrt{q_{n+1}^2 - 1}\right) \\
&= \frac{1}{2q_n}\left(2q_n^2 - \sqrt{4q_n^4 - 4q_n^2}\right) \\
&= \frac{1}{2q_n}\left(2q_n^2 - 2q_n\sqrt{q_n^2 - 1}\right) \\
&= q_n - \sqrt{q_n^2 - 1}. \ \square
\end{aligned}$$

Theorem 24: If $q_0 \geq 1$ and (23.1) holds, then

$$q_0 - \sqrt{q_0^2 - 1} = \sum_{n=0}^{\infty} \frac{1}{2^{n+1} q_0 q_1 \cdots q_n}$$

is an Engel series.

Proof: Note that by the lemma, we have

$$\begin{aligned}
q_0 - \sqrt{q_0^2 - 1} &= \frac{1}{2q_0} + \frac{1}{2q_0}\left(q_1 - \sqrt{q_1^2 - 1}\right) \\
&= \frac{1}{2q_0} + \frac{1}{2q_0 q_1} + \frac{1}{2q_1}\left(q_2 - \sqrt{q_2^2 - 1}\right) \\
&= \cdots \\
&= \sum_{m=0}^{n} \frac{1}{2^{m+1} q_0 \cdots q_m} + q_{n+1} - \sqrt{q_{n+1}^2 - 1}. \qquad (24.1)
\end{aligned}$$

Since $q_n \to \infty$ rapidly, the remainder goes to zero very rapidly. $\square$

Chapter 5

Sylvester Series

5.1 Construction

As with the Engel series, we start with an $x_0 \in (0, 1]$ and define:

$$\alpha_0 = \frac{1}{x_0} \in [1, \infty), \qquad q_0 = [\alpha_0] + 1 = \left[\frac{1}{x_0}\right] + 1. \tag{25.1}$$

Note that q_0 is identical to what was the start of the Engel series. Now we make a change. Define α_1 by

$$\frac{1}{\alpha_0} = \frac{1}{q_0} + \frac{1}{\alpha_1}. \tag{25.2}$$

From (25.1), we have that $\alpha_0 < q_0$. Note that

$$\frac{1}{\alpha_1} = \frac{1}{\alpha_0} - \frac{1}{q_0} = \frac{q_0 - \alpha_0}{\alpha_0} = \frac{[\alpha_0] + 1 - \alpha_0}{\alpha_0([\alpha_0] + 1)} \leq \frac{1}{\alpha_0([\alpha_0] + 1)} < \frac{1}{\alpha_0^2}. \tag{25.3}$$

Now define iteratively

$$q_n = [\alpha_n] + 1, \qquad \frac{1}{\alpha_n} = \frac{1}{q_n} + \frac{1}{\alpha_{n+1}}. \tag{25.4}$$

We then get that

$$\alpha_{n+1} > \alpha_n^2. \tag{25.5}$$

So, then the α_n are increasing rapidly. Also,

$$\frac{1}{\alpha_{n+1}} = \frac{[\alpha_n] + 1 - \alpha_n}{\alpha_n([\alpha_n] + 1)} < \frac{1}{\alpha_n^2}.$$

Now, we have that

$$1 \leq \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_n < \cdots \to \infty, \text{ and}$$
$$2 \leq q_0 \leq q_1 \leq \cdots \to \infty. \tag{25.6}$$

Iterating (25.4), we get the following:

$$x_0 = \frac{1}{\alpha_0} = \frac{1}{q_0} + \frac{1}{\alpha_1} = \frac{1}{q_0} + \frac{1}{q_1} + \frac{1}{\alpha_2} = \cdots = \sum_{m=0}^{n} \frac{1}{q_m} + \frac{1}{\alpha_{n+1}}, \tag{25.7}$$

and thus, letting $n \to \infty$, we have

$$x_0 = \sum_{m=0}^{\infty} \frac{1}{q_m}. \tag{25.8}$$

Remark: We can find a lower bound for q_m. From the definition of q_m in (25.4), we have that

$$q_n - 2 \leq \alpha_n < q_n.$$

Thus,

$$\frac{1}{q_n - 1} \geq \frac{1}{\alpha_n} = \frac{1}{q_n} + \frac{1}{\alpha_{n+1}} > \frac{1}{q_n} + \frac{1}{q_{n+1}}. \tag{25.9}$$

Therefore,

$$\frac{1}{q_{n+1}} < \frac{1}{q_n - 1} - \frac{1}{q_n} = \frac{1}{q_n(q_n) - 1} \Longrightarrow q_{n+1} > q_n^2 - q_n. \tag{25.10}$$

Thus,

$$q_{n+1} \geq q_n^2 - q_n + 1. \tag{25.11}$$

5.2 Some Theorems

Theorem 25-1: Every $x_0 \in (0,1]$ can be represented as a series

$$x_0 = \sum_{m=0}^{\infty} \frac{1}{q_m},$$

with $q_m \to \infty$ and q_m satisfying $q_{n+1} \geq q_n^2 - q_n + 1$.

Proof: See the above discussion. □

Theorem 25-2: If $x_0 \in (0,1]$ is given by (25.8) and the $q_n \to \infty$ satisfy (25.11), then q_n are uniquely defined by (25.4).

Proof: Define $\alpha_0 := \frac{1}{x_0}$ such that $a_0 \in [1, \infty)$. Define α_1 by

$$\frac{1}{\alpha_0} = \frac{1}{q_0} + \frac{1}{\alpha_1}.$$

Then, $\alpha_1 = \sum_{m=1}^{\infty} \frac{1}{q_m}$.

Clearly, $x_0 = \frac{1}{\alpha_0} > \frac{1}{q_0} \Longrightarrow q_0 > \alpha_0$. The inequality (25.11) implies that

$$q_{n+1} - 1 \geq q_n(q_n - 1).$$

Therefore,

$$\frac{1}{q_{n+1} - 1} \leq \frac{1}{q_n(q_n - 1)} = \frac{1}{q_n - 1} - \frac{1}{q_n}.$$

Hence,

$$\frac{1}{q_n} \leq \frac{q}{q_n - 1} - \frac{1}{q_{n+1} - 1}, \tag{25.12}$$

and the right hand side telescopes.

Therefore, $x_0 = \frac{1}{\alpha_0} = \sum_{n=0}^{\infty} \frac{1}{q_n} \leq \sum_{n=0}^{\infty} \left(\frac{1}{q_n - 1} - \frac{1}{q_{n+1} - 1} \right) = \frac{1}{q_0 - 1}$.

Thus, $\alpha_0 \geq q_0 - 1$, and so $q_0 = [\alpha_0] + 1$, and then we iterate. □

5.3 An Irrationality Criterion

Theorem 26: Let $x_0 \in (0,1]$ be represented by the Sylvester series

$$x_0 = \sum_{m=1}^{\infty} \frac{1}{q_m},$$

where $q_m \in \mathbb{Z}^+$, $q_0 \geq 2$, and $q_{n+1} \geq q_n^2 - q_n + 1$, for all $n \geq 0$. Then, $x_0 \in \mathbb{Q}$ if and only if there exists $N \in \mathbb{Z}^+$ such that

$$q_{n+1} = q_n^2 - q_n + 1 \tag{26.1}$$

for all $n \geq N$.

Proof: ($\Longleftarrow$) If (26.1) holds, then

$$\frac{1}{q_{n+1}-1} = \frac{1}{q_n^2 - q_n} = \frac{1}{q_n(q_n-1)} = \frac{1}{q_n - 1} - \frac{1}{q_n}, \text{ for all } n \geq N,$$

and so

$$\frac{1}{q_n} = \frac{1}{q_n - 1} - \frac{1}{q_{n+1} - 1},$$

which has a telescoping property. Thus,

$$\begin{aligned} x_0 &= \sum_{m=0}^{\infty} \frac{1}{q_m} \\ &= \sum_{m=0}^{N-1} \frac{1}{q_m} + \sum_{m=N}^{\infty} \frac{1}{q_m} \\ &= \sum_{m=0}^{N-1} \frac{1}{q_m} + \sum_{m=N}^{\infty} \left(\frac{1}{q_m - 1} - \frac{1}{q_{m+1} - 1} \right) \\ &= \sum_{m=0}^{N-1} \frac{1}{q_m} + \frac{1}{q_N - 1} \text{ (by telescoping), which is rational.} \end{aligned}$$

($\Longrightarrow$) Let $x_0 \in (0,1]$, with $x_0 \in \mathbb{Q}$. Write

$$x_0 = \frac{p}{q}, \text{ with } p, q \in \mathbb{Z}^+, 1 \leq p \leq q, \text{ and } \gcd(p,q) = 1.$$

Recall (25.7):

$$\frac{1}{\alpha_n} = \sum_{m=n}^{\infty} \frac{1}{q_m},$$

and so,

$$\frac{1}{\alpha_n} = x_0 - \sum_{m=0}^{n-1} \frac{1}{q_m} = \frac{k_n}{q \cdot q_0 \cdots q_{n-1}}, \tag{26.2}$$

for some $k_n \in \mathbb{Z}^+$. Also, since

$$\frac{1}{\alpha_n} = \frac{1}{q_n} + \frac{1}{\alpha_{n+1}},$$

we have

$$\frac{k_n}{q \cdot q_0 \cdots q_{n-1}} = \frac{1}{q_n} + \frac{k_{n+1}}{q \cdot q_0 \cdots q_n} \Longleftrightarrow k_n q_n = q q_0 \cdots q_{n-1} + k_{n+1}. \tag{26.3}$$

Recall (25.9):

$$q_n - 1 \leq \alpha_n < q_n,$$

and thus

$$\alpha_{n+1} \geq q_{n+1} - 1 \geq q_n^2 - q_n. \tag{26.4}$$

But by definition and (26.4),

$$\alpha_{n+1} = \frac{qq_0 \cdots q_n}{k_{n+1}} \geq q_n^2 - q_n.$$

Canceling q_n, we have that

$$qq_0 \cdots q_{n-1} \geq k(q_n - 1). \tag{26.5}$$

Combining (26.5) and (26.3), we get that

$$k_n q_n \geq k_{n+1}(q_n - 1) + k_{n+1} = k_{n+1} q_n.$$

So, the k_n is a decreasing sequence of positive integers, and thus are constant from some point on. Say that $k_n = k$ for all $n \geq N$. Therefore,

$$\frac{1}{\alpha_{n+1}} = \frac{k}{qq_0 \cdots q_n} = \frac{1}{q_n \alpha_n} = \frac{1}{q_n}\left(\frac{1}{q_n} + \frac{1}{\alpha_{n+1}}\right). \tag{26.6}$$

Thus, for all $n \geq N$,

$$\frac{1}{\alpha_{n+1}}\left(1 - \frac{1}{q_n}\right) = \frac{1}{q_n^2} \Longleftrightarrow \alpha_{n+1} = q_n(q_n - 1) = q_n^2 - q_n \in \mathbb{Z}. \tag{26.7}$$

Recall that $q_n = [\alpha_n] + 1$, from (25.4). So, $q_{n+1} = [\alpha_{n+1}] + 1 = q_n^2 - q_n + 1$, for all $n \geq N$. □

Remarks:

(1) The Engel series for $x_0 \in (0, 1]$ and the Sylvester series for $x_0 \in (0, 1]$ both start identically.

Engel: $u_0 = \left\lceil \frac{1}{x_0} \right\rceil + 1 \geq 2$.
Sylvester: $\alpha_0 = \frac{1}{x_0}, \quad q_0 = [\alpha_0] + 1$.

So, at the first step, $q_0 = u_0$. The difference takes place in the next step.

Engel: Define $x_1 := x_1 = u_0 x_0 - 1 \Leftrightarrow x_0 = \frac{1}{u_0} + \frac{x_1}{u_0}$.
Sylvester: $\frac{1}{\alpha_0} = x_0 = \frac{1}{q_0} + \frac{1}{\alpha_2}$, so $\alpha_1 = \frac{u_0}{x_1}$.

So, in the Sylvester case, we don't require the divisibility condition required by the Engel case.

Example: Let $x_0 = 1$. In Engel, $u_0 = 2$, so $x_1 = 1$, and thus $u_n = 2$ for all n. Thus the Engel series for 1 is

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \cdots$$

In Sylvester, $q_0 = 2$, and $q_1 = 2^2 - 2 + 1 = 3$, and $q_2 = 3^2 - 3 + 1 = 7$, and $q_3 = 7^2 - 7 + 1 = 43$. Thus the Sylvester series for 1 is

$$1 = \frac{1}{2} + \frac{1}{3} + \frac{1}{7} + \frac{1}{43} + \cdots$$

(2) Sylvester has gotten rid of the divisibility condition essential for Engel and replaced it with a growth condition. Thus, the Sylvester series and criterion is more widely applicable. This yields:

Theorem 27: For $a \in \mathbb{Z}^+$, define the generalized Fermat numbers by

$$F_{n,a} = 2^{2^n} - a, \qquad (\bigstar)$$

such that $F_{n,a} \neq 0$. Then,

$$\sum \frac{1}{F_{n,a}} \in J \text{ is irrational.}$$

Proof: Note that $F_{n+1,a} = 2^{2^{n+1}} - a \geq F_{n,a}^2 - F_{n,a} + 1 = (2^{2^n} - a)^2 - (2^{2^n} - a) + 1 = 2^{2^{n+1}} - 2a2^{2^n} + a^2 + a + 1$, with the last three terms being fixed. Thus, this sum is irrational by the Sylvester Condition. □

(3) When $a = 1$, the numbers $2^{2^n} - 1$ satisfy the Divisibility conditions (Engel) and the truncations yield strong approximations in the sense of Dirichlet, whereas when $a > 0$ for a odd, the truncations do not produce strong approximations because

$$\text{lcm}\{F_{m,a}\}_{m=0}^{m=n} > F_{n+1,a}, \text{ in fact, } \textit{much} \text{ larger.}$$

Chapter 6

Continued Fractions

6.1 General Continued Fractions

Remark: Our motivation is to find an iterative process which works in general (not just for specific series) that encapsulates the strong approximations of Dirichlet.

Definition: By a general continued fraction, we mean an expression of the following type:

$$R_n = a_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{\ddots \cfrac{}{b_{n-1} + \cfrac{a_n}{b_n}}}}}, \tag{28.1}$$

where $a_m, b_m \in \mathbb{C}$. We also will consider the infinite expression:

$$R = a_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \ddots}}}. \tag{28.2}$$

For simplicity, we write

$$R_n = a_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\frac{a_3}{b_3+}\cdots\frac{a_{n-1}}{b_{n-1}+}\frac{a_n}{b_n}. \tag{28.3}$$

and

$$R = a_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\frac{a_3}{b_3+}\cdots. \tag{28.4}$$

Consider the sequence of truncations of R_n given by

$$R_0 := a_0 = \frac{a_0}{1}, \qquad R_1 := a_0 + \frac{a_1}{b_1} = \frac{a_0 b_1 + a_1}{b_1}, \ldots. \tag{28.5}$$

Thus, for each R_m, we may define P_m and Q_m to be the m^{th} numerator and denominator, respectively, calculated as they occur, without any cancellation.With this notation, $P_0 = a_0$, $P_1 = a_0b_1 + a_1$, $Q_0 = 1$, and $Q_1 = b_1$.

Theorem 28: Define $P_{-1} = 1$ and $Q_{-1} = 0$. Define $P_0 = a_0$ and $Q_0 = 1$. Define P_m and Q_m recursively by:

$$\left\{ \begin{array}{c} P_m = b_m P_{m-1} + a_m P_{m-2} \\ Q_m = b_m Q_{m-1} + a_m Q_{m-2} \end{array} \right\}, \tag{28.6}$$

for $m \geq 1$. Then, P_m and Q_m are the m^{th} numerator and m^{th} denominator of R_n (or R) respectively.

Proof: (by induction) From (28.5), we see that (28.6) holds for $m = 1$, and so for all fractions of length 1. Assume that (28.6) holds for all fractions of length m.

We write that

$$a_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\cdots\frac{a_m}{b_m+}\frac{a_{m+1}}{b_{m+1}} = a_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\cdots\frac{a_{m-1}}{b_{m-1}+}\frac{b_{m+1}a_m}{b_m b_{m+1} + a_{m+1}}. \tag{28.7}$$

We note that this converts a fraction of length $m+1$ into a fraction of length m.

Thus P_{m+1} and Q_{m+1} may be calculated from (28.6) by replacing

$$a_m \longmapsto b_{m+1}a_m, \qquad b_m \longmapsto b_m b_{m+1} + a_{m+1}. \tag{28.8}$$

Thus,

$$\begin{aligned} P_m &= b_m b_{m+1} P_{m-1} + a_{m+1} P_{m-1} + b_{m+1} a_m P_{m-2} \\ &= b_{m+1}(b_m P_{m-1} + a_m P_{m-2}) + a_{m+1} P_{m-1} \\ &= b_{m+1} P_m + a_{m+1} P_{m_1}, \text{ (by the induction hypothesis)} \end{aligned} \tag{28.9}$$

which is (28.6) for $m+1$. Similar for Q_m. □

With P_n, Q_n determined by **Theorem 28**, we define the value of R_n to be $R_n = \frac{P_n}{Q_n}$, provided $Q_n \neq 0$. In the case that $Q_n = 0$, we write *formally* that $R_n = \frac{P_n}{0}$.

Example: Consider $R_3 := 1 + \dfrac{1}{1 - \dfrac{1}{1 + \dfrac{1}{1}}}$. Now,

$$R_0 = \frac{1}{1}, \quad R_1 = 1 + \frac{1}{1} = \frac{2}{1}, \quad R_2 = 1 + \frac{1}{1 - \frac{1}{1}} = 1 + \frac{1}{0} = \frac{1}{0}, \quad \text{but } R_3 = 1 + \frac{1}{1 - \frac{1}{2}} = 1 + \frac{2}{1} = \frac{3}{1}.$$

6.2 Convergence of General Continued Fractions

We say that the infinite continued fraction R in (28.2) converges if

$$\lim_{n\to\infty} \frac{P_n}{Q_n} = \ell < \infty$$

exists with P_n, Q_n defined by **Theorem 28**.

Theorem 29-1: For $n \geq 0$, the numerators and denominators satisfy

$$P_n Q_{n-1} - P_{n-1} Q_n = (-1)^{n-1} a_1 a_2 \cdots a_n. \tag{29.1}$$

Proof: (by induction) The theorem is true for $n = 1$ because $P_1 Q_0 - P_0 Q_1 = (a_0 b_1 + a_1) \cdot 1 - a_0 b_1 = a_1$. Note that (29.1) holds also for $n = 0$, since for $n = 0$, the product $a_1 a_2 \cdot a_n$ is a null product.

Assume (29.1) holds for a certain m. Then,

$$\begin{aligned} P_{m+1} Q_m - P_m Q_{m+1} &= Q_m (b_{m+1} P_m + a_{m+1} P_{m-1}) - P_m (b_{m+1} Q_m + a_{m+1} Q_{m-1}) \\ &= -a_{m+1} (P_m Q_{m-1} - P_{m-1} Q_m) \\ &= -a_{m+1} (-1)^{m-1} a_1 a_2 \cdots a_m, \text{ (by Induction Hypothesis)} \\ &= (-1)^m a_1 a_2 \cdots a_m. \end{aligned} \tag{29.2}$$

Hence, true. □

Theorem 29-2: If $Q_n \neq 0$ and $Q_{n-1} \neq 0$ for some n, then

$$R_n - R_{n-1} = \frac{(-1)^{n-1} a_1 \cdots a_n}{Q_n Q_{n-1}}.$$

Proof: Since $Q_n \neq 0$ and $Q_{n-1} \neq 0$, we have

$$R_n = \frac{P_n}{Q_n}, \text{ and, } R_{n-1} = \frac{P_{n-1}}{Q_{n-1}}.$$

So,

$$R_n - R_{n-1} = \frac{P_n}{Q_n} - \frac{P_{n-1}}{Q_{n-1}} = \frac{P_n Q_{n-1} - P_{n-1} Q_n}{Q_n Q_{n-1}} = \frac{(-1)^{n-1} a_1 \cdots a_n}{Q_n Q_{n-1}}. \ \square$$

Theorem 30: If $\{Q_m\}$ is zero-free, then the infinite continued fraction R in (28.2) converges if and only if

$$\sum \frac{(-1)^{n-1} a_1 \cdots a_n}{Q_n Q_{n-1}} \text{ is convergent.} \tag{30.1}$$

If so,

$$R = a_0 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} a_1 \cdots a_n}{Q_n Q_{n-1}}.$$

Proof: Note that

$$\begin{aligned} \sum_{n=1}^{N} (R_n - R_{n-1}) &= R_N - R_0, \text{ by telescoping} \\ &= \sum_{n=1}^{N} \frac{(-1)^{n-1} a_1 \cdots a_n}{Q_n Q_{n-1}}. \end{aligned} \tag{30.2}$$

Now, R converges if and only if $\lim R_N = \ell < \infty$ exists, if and only if the series in (30.1) converged. □

For ease of application, we prove the next result.

Theorem 31: If we have

$$\sum_{n=1}^{\infty} \frac{a_1 a_2 \cdots a_n}{b_1 b_2 \cdots b_n} =: \ell < \infty, \tag{31.1}$$

and if $a_n + b_n \neq 0$, for all $n \geq 1$, then

$$R = \frac{a_1}{b_1 -} \frac{a_2 b_1}{(a_2 + b_2) -} \frac{a_3 b_2}{(a_3 + b_3) -} \cdots \tag{31.2}$$

converges to the value ℓ given by (31.1).

Proof: We may set $a_0 = 0$. Calculate the following:

$$R_1 = \frac{a_1}{b_1} \implies P_1 = a_1,\ Q_1 = b_1. \tag{31.3}$$

$$R_2 = \frac{a_1}{b_1 - \dfrac{a_2 b_1}{a_2 + b_2}} = \frac{a_1(a_2 + b_2)}{b_1(a_2 + b_2) - a_2 b_1} = \frac{a_1 a_2 + a_1 b_2}{b_1 b_2} = \frac{a_1}{b_1} + \frac{a_1 a_2}{b_1 b_2}. \tag{31.4}$$

We now make two claims:

Claim 1: $Q_n = b_1 \cdots b_n$, hence Q_n is zero-free.

Claim 2: $R_n = \displaystyle\sum_{m=1}^{n} \frac{a_1 \cdots a_m}{b_1 \cdots b_m}$.

Proof of Claim 1: (by induction)

True for $n = 1, 2$. Assume true for all $m \leq n$. Then

$$\left.\begin{array}{rcl} Q_{n+1} & = & (a_{n+1} + b_{n+1}) Q_n - a_{n+1} b_n Q_{n-1} \\ & = & (a_{n+1} + b_{n+1})(b_1 \cdots b_n) - a_{n+1} b_n (b_1 \cdots b_{n-1}) \\ & = & b_1 \cdots b_{n+1}. \end{array}\right\} \tag{31.5}$$

Proof of Claim 2:

To prove this, we only need to show that

$$R_n - R_{n-1} = \frac{a_1 \cdots a_n}{b_1 \cdots b_n}. \tag{31.6}$$

By **Theorem 29.ii**, and (31.2) we get that

$$R_n - R_{n-1} = \frac{(-1)^{n-1}(a_1)(a_2 b_1)(a_3 b_2) \cdots (a_n b_{n-1})(-1)^{n-1}}{Q_n Q_{n-1}} = \frac{a_1 \cdots a_n b_1 \cdots b_{n-1}}{b_1 \cdots b_n b_1 \cdots b_{n-1}} = \frac{a_1 \cdots a_n}{b_1 \cdots b_n},$$

which is (31.6).

The theorem follows. □

6.3 Continued Fractions for e and π

Theorem 32: We have the following contented fraction for e:

Proof: Note that

$$e^{-1} = \frac{1}{2} - \frac{1}{2 \cdot 3} + \frac{1}{2 \cdot 3 \cdot 4} - \cdots .$$

We may write this as

$$\sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{b_1 \cdots b_n},$$

with $a_1 = 1$ and $a_i = -1$ for $i \geq 2$, and $b_i = i + 1$ for $i \geq 1$.

So, we get that

$$e^{-1} = \cfrac{1}{2 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \ddots}}}}.$$

Therefore,

$$e = 2 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \ddots}}}. \ \square$$

Theorem 33: (Lord Brouncher) We have the following continued fraction for π:

Proof: To get a continued fraction for π, we consider the (Newton / Gregory / Leibniz) Series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots .$$

This can be written as

$$\sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{b_1 \cdots b_n},$$

with $a_1 = 1$, $a_n = -(2n - 3)$, $b_1 = 1$, $b_n = 2n - 1$, for $n \geq 2$. So,

$$\frac{\pi}{4} = \cfrac{1}{1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{\ddots}}}}}. \ \square$$

Comparison of Theorems 32 and 33.

(1) The truncations of the fraction for e^{-1} yield approximations p_n/q_n good enough to satisfy the Dirichlet Criterion that $0 \neq |q_n e^{-1} - p_n| \to 0$, whereas the truncations of the fraction for $\frac{\pi}{4}$ **do not** satisfy $0 \neq |q_n \frac{\pi}{4} - p_n| \to 0$.

(2) The truncations of the fraction for e^{-1} **are not** strong approximations in the sense that $|q_n e^{-1} - p_n| < \frac{1}{q_n}$. Such strong approximations will be produced by standard / simple continued fractions.

History: Lord Brouncher found **Theorem 33** by recasting a famous formula due to Wallace:

$$\frac{\pi}{2} = \frac{2 \cdot 2}{1 \cdot 3} \cdot \frac{4 \cdot 4}{3 \cdot 5} \cdot \frac{6 \cdot 6}{5 \cdot 7} \cdots .$$

The method of proof given above using the Newton series was due to Euler.

6.4 Conversion of Continued Fractions

Definition: A regular continued fraction is one for which $a_i = 1$ for $i \geq 1$, i.e.,

$$R = a_0 + \frac{1}{c_1+}\frac{1}{c_2+}\frac{1}{c_3+}\cdots .$$

Note that

$$\frac{a_1}{b_1 + \dfrac{a_2}{b_2 + \dfrac{a_3}{\ddots}}} = \frac{1}{\dfrac{b_1}{a_1} + \dfrac{a_2/a_1}{b_2 + \dfrac{a_3}{\ddots}}} = \frac{1}{\dfrac{b_1}{a_1} + \dfrac{1}{\dfrac{a_1b_2}{a_2} + \dfrac{a_1a_3/a_2}{b_3 + \dfrac{a_4}{\ddots}}}} = \frac{1}{\dfrac{b_1}{a_1} + \dfrac{1}{\dfrac{a_1b_2}{a_2} + \dfrac{1}{\dfrac{a_2b_3}{a_1a_3} + \dfrac{1}{\ddots}}}}.$$

Define

$$c_1 = \frac{b_1}{a_1}, \quad c_2 = \frac{a_1b_2}{a_3}, \quad c_3 = \frac{a_2b_3}{a_1a_3}, \ldots$$

Theorem 34: If $\{a_n\}$, $\{b_n\}$ are two infinite sequences of non-zero complex numbers, then

$$\frac{a_1}{b_1+}\frac{a_2}{b_2+}\frac{a_3}{b_3+}\cdots = \frac{1}{c_1+}\frac{1}{c_2+}\frac{1}{c_3+}\cdots ,$$

where

$$c_{2n} = \frac{a_1a_3\cdots a_{2n-1}b_{2n}}{a_2a_4\cdots a_{2n}}, \text{ and}$$

$$c_{2n+1} = \frac{a_2a_4\cdots a_{2n}b_{2n+1}}{a_1a_3\cdots a_{2n+1}}.$$

(Since we're not talking about convergence, what we're really saying is that for any truncation, the two sides are equal.)

Proof: Homework. Use induction on n.

Theorem 35: Let $\{c_n\}$ be an infinite sequence of complex numbers such that

$$|c_n| \geq 2, \text{ for all } n \geq 1, \text{ and} \tag{35.1}$$

$$\sum_{n=1}^{\infty} \frac{1}{|c_nc_{n+1}|} = \ell < \infty. \tag{35.2}$$

Then, the regular continued fraction

$$\frac{1}{c_1+}\frac{1}{c_2+}\cdots , \tag{35.3}$$

is convergent.

Proof: The numerators P_n and denominators Q_n of R satisfy the following recurrences:

$$\left.\begin{aligned} P_n &= c_nP_{n-1} + P_{n-2}, \\ Q_n &= c_nQ_{n-}1 + Q_{n-2}, \end{aligned}\right\} \tag{35.4}$$

for $n \geq 2$, with

$$P_0 := 0, \quad Q_0 = 1, \quad P_1 = 1, \quad Q_1 = c_1.$$

We claim that

$$|P_{n-1}| \leq |P_n| \text{ and } |Q_{n-1}| \leq |Q_n|, \text{ for all } n \in \mathbb{N}. \tag{35.5}$$

Once we prove this, it follows that P_n and Q_n are nonzero for $n > 0$.

To see this, note that (35.4) and (35.1) imply that

$$|P_n| \geq ||c_n P_{n-1}| - |P_{n-2}|| \geq 2|P_{n-1}| - |P_{n-2}| \geq 2|P_{n-1}| - |P_{n-1}| = |P_{n-1}|,$$

inductively. Similarly for Q_n. In particular, P_n and Q_n are zero-free.

Now, define two sequences u_n, v_n by

$$u_n := c_n \frac{P_{n-1}}{P_n}, \qquad v_n := c_n \frac{Q_{n-1}}{Q_n}. \tag{35.6}$$

Therefore,

$$P_{n-2} = \frac{u_{n-1} P_{n-1}}{c_{n-1}}, \quad \text{and} \quad Q_{n-2} = \frac{v_{n-1} Q_{n-1}}{c_{n-1}}. \tag{35.7}$$

Hence (35.7) and (35.4) yield

$$P_n = c_n P_{n-1} + \frac{u_{n-1} P_{n-1}}{c_{n-1}} = c_n P_{n-1} \left(1 + \frac{u_{n-1}}{c_{n-1} c_n}\right). \tag{35.8a}$$

Similarly,

$$Q_n = c_n Q_{n-1} \left(1 + \frac{v_{n-1}}{c_{n-1} c_n}\right). \tag{35.8b}$$

By iteration of (35.8), we get that

$$P_n = c_2 c_3 \cdots c_n \prod_{m=1}^{n-1} \left(1 + \frac{u_m}{c_m c_{m+1}}\right), \tag{35.9a}$$

$$Q_n = c_1 c_2 c_3 \cdots c_n \prod_{m=1}^{n-1} \left(1 + \frac{v_m}{c_m c_{m+1}}\right). \tag{35.9b}$$

Note also that (35.6) combined with (35.8) yield

$$u_n = \frac{1}{1 + \dfrac{u_{n-1}}{c_{n-1} c_n}}, \qquad \text{and} \qquad v_n = \frac{1}{1 + \dfrac{v_{n-1}}{c_{n-1} c_n}}. \tag{35.10}$$

Since $u_0 = 0$ and $v_0 = 1$, we have that $|u_1|, |v_1| \leq 2$. Thus (35.10) implies inductively that

$$|u_n| \leq 2 \text{ and } |v_n| \leq 2, \tag{35.11}$$

since

$$|u_n| \leq \frac{1}{1 - \left|\dfrac{u_{n-1}}{c_{n-1} c_n}\right|} \leq \frac{1}{1 - \dfrac{2}{2 \cdot 2}} = 2.$$

This implies that

$$\left.\begin{aligned} &\sum \left|\frac{u_m}{c_m c_{m+1}}\right| \text{ is convergent, and} \\ &\sum \left|\frac{v_m}{c_m c_{m+1}}\right| \text{ is convergent.} \end{aligned}\right\} \tag{35.12}$$

This means that the infinite products

$$\alpha := \prod_{m=1}^{\infty} \left(1 + \frac{u_m}{c_m c_{m+1}}\right), \quad \text{and} \quad \beta := \prod_{m=1}^{\infty} \left(1 + \frac{v_m}{c_m c_{m+1}}\right), \tag{35.13}$$

are convergent. Thus (35.9) implies that

$$\lim_{n\to\infty} \frac{P_n}{Q_n} = \frac{\alpha}{c_1 \beta}$$

exists. Thus the infinite continued fraction is convergent. □

Remark: For convergence, it suffices to have $|c_n| \geq 2$, for all $n \in \mathbb{N}$. Additionally, convergent infinite products are non-zero in value. (By definition we say that if an infinite product equals zero, that it **diverges** to zero.)

6.5 Bessel Functions: Their Quotients and Their Irrationals

First, we need to define the Gamma function.

Definition:

$$\Gamma(s) := \int_0^\infty t^{s-1}e^{-t}dt, \tag{36.1}$$

for $\Re(s) > 0$. It satisfies the functional equation

$$\Gamma(s+1) = s\Gamma(s). \tag{36.2}$$

This functional equation extends Γ to be a meromorphic function on $\mathbb{C}$ with simple poles only at 0, -1, -2, etc, and elsewhere analytic. Also, $\Gamma(n+1) = n!$, for $n \in \mathbb{N}$.

Remark: A more striking value is:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{36.3}$$

To realize this, observe:

$$\begin{aligned}
\Gamma\left(\frac{1}{2}\right) &= \int_0^\infty t^{-\frac{1}{2}}e^{-t}dt \\
&= \int_0^\infty 2e^{-u^2}du \\
&= \int_{-\infty}^\infty e^{-u^2}\,du \\
&= \sqrt{\int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(u^2+v^2)}\,du\,dv} \\
&= \sqrt{\int_0^\infty \int_0^{2\pi} e^{-r^2}r\,dr\,d\theta} \\
&= \sqrt{\left(\int_0^\infty e^{-r^2}r\,dr\right)\left(\int_0^{2\pi} d\theta\right)} \\
&= \sqrt{\frac{1}{2}\cdot 2\pi} \\
&= \sqrt{\pi}.
\end{aligned}$$

Definition: (Bessel Function) Let $x \in \mathbb{C}$. Then,

$$J_v(x) := \sum_{k=0}^\infty \frac{(-1)^k}{k!\cdot\Gamma(v+k+1)}\left(\frac{x}{2}\right)^{2k+v}, \tag{36.4}$$

for $v > -1$ real. Clearly the series is convergent for all $x \in \mathbb{C}$. It is well defined provided that x^v is well defined. We use the formulas

$$\left.\begin{aligned} J_{\frac{1}{2}}(x) &= \sqrt{\frac{2}{\pi}}x^{-\frac{1}{2}}\sin x \\ J_{-\frac{1}{2}}(x) &= \sqrt{\frac{2}{\pi}}x^{-\frac{1}{2}}\cos x \end{aligned}\right\} \tag{36.5}$$

To see this:

$$\begin{aligned}
J_{\frac{1}{2}}(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+\frac{1}{2})(k-\frac{1}{2})\cdots\frac{1}{2}\cdot\Gamma(\frac{1}{2})}\left(\frac{x^{2k}\sqrt{x}}{2^{2k}\sqrt{2}}\right) \\
&= \sqrt{\frac{2}{\pi}}x^{-\frac{1}{2}}\sum_{k=0}^{\infty}\frac{(-1)^k\cdot x^{2k+1}}{k!(k+\frac{1}{2})(k-\frac{1}{2})\cdots\frac{1}{2}\cdot 2^{2k+1}} \\
&= \sqrt{\frac{2}{\pi}}x^{-\frac{1}{2}}\sum_{k=0}^{\infty}\frac{(-1)^k x^{2k+1}}{(2k+1)!} \\
&= \sqrt{\frac{2}{\pi}}x^{-\frac{1}{2}}\sin x.
\end{aligned}$$

Similarly for $J_{-\frac{1}{2}}$.

Consequently,

$$\frac{J_{\frac{1}{2}}(x)}{J_{-\frac{1}{2}}(x)} = \tan x, \tag{36.6}$$

when $J_{-\frac{1}{2}}(x) \neq 0$.

A recurrence for $J_\nu(x)$:

The $J_\nu(x)$ satisfy

$$J_{\nu-1}(x) + J_{\nu+1}(x) = \frac{2\nu}{x}J_\nu(x), \tag{36.7}$$

if $x \neq 0$ and $\nu > 0$. To see (36.7), note that

$$\begin{aligned}
J_{\nu-1}(x) + J_{\nu+1}(x) &= \left(\frac{x}{2}\right)^{\nu-1}\sum_{n=0}^{\infty}\frac{(-x^2/4)^n}{n!\,\Gamma(\nu+n)} + \left(\frac{x}{2}\right)^{\nu+1}\sum_{n=0}^{\infty}\frac{(-x^2/4)^n}{n!\,\Gamma(\nu+n+2)}. \\
&= \left(\frac{x}{2}\right)^{\nu-1}\left[\sum_{n=0}^{\infty}\frac{(-x^2/4)^n}{n!\,\Gamma(\nu+n)} - \sum_{n=0}^{\infty}\frac{(-x^2/4)^{n+1}}{n!\,\Gamma(\nu+n+2)}\right] \\
&= \left(\frac{x}{2}\right)^{\nu-1}\left[\frac{1}{\Gamma(v)} + \sum_{n=1}^{\infty}\frac{(-x^2/4)^n}{n!\,\Gamma(\nu+n)} - \frac{(-x^2/4)^n}{(n-1)!\,\Gamma(\nu+n+1)}\right] \\
&= \left(\frac{x}{2}\right)^{\nu-1}\left[\frac{\nu}{\Gamma(\nu+1)} + \sum_{n=1}^{\infty}\frac{(-x^2/4)^n\cdot(\nu+n-n)}{n!\,\Gamma(\nu+n+1)}\right] \\
&= \left(\frac{x}{2}\right)^{\nu-1}\frac{\nu}{\Gamma(\nu+1)} + \sum_{n=1}^{\infty}\frac{(-x^2/4)^n\cdot\nu}{n!\,\Gamma(\nu+n+1)} \\
&= \left(\frac{x}{2}\right)^{\nu}\frac{2\nu}{x}\sum_{n=0}^{\infty}\frac{(-1)^n(x/2)^{2n}}{n!\,\Gamma(\nu+n+1)} \\
&= J_\nu(x)\frac{2\nu}{x}
\end{aligned}$$

Asymptotic Behavior of $J_\nu(x)$ as $\nu \to \infty$:

Fix x and let $\nu \to \infty$. Then,

$$J_\nu(x) = \frac{(x/2)^\nu}{\Gamma(\nu+1)}\left[1 + O\left(\frac{1}{\nu}\right)\right],$$

where "O()" is big-oh notation.

Hence,

$$J_\nu(x) \sim \frac{(x/2)^\nu}{\Gamma(\nu+1)}. \tag{36.8}$$

To set up a continued fraction, rewrite the recurrence as

$$J_{\nu-1}(x) = \frac{2\nu}{x} J_\nu(x) - J_{\nu+1}(x), \text{ and}$$

$$\frac{J_{\nu-1}(x)}{J_\nu(x)} = \frac{2\nu}{x} - \frac{J_{\nu+1}(x)}{J_\nu(x)},$$

provided that $J_\nu(x) \neq 0$.

Hence,

$$\frac{J_\nu(x)}{J_{\nu-1}(x)} = \frac{1}{\frac{2\nu}{x} - \frac{J_{\nu+1}(x)}{J_\nu(x)}}, \tag{36.9}$$

provided that $J_{\nu-1}(x)$ is also non-zero. Since J_ν is an analytic function, it's zero-set is countable and will not cluster. Thus, the zero set of $J_\nu J_{\nu-1}$ has the same property.

Iterating (36.9), we get that

$$\begin{aligned}\frac{J_\nu(x)}{J_{\nu-1}(x)} &= \frac{1}{\frac{2\nu}{x} - \frac{1}{\frac{J_\nu(x)}{J_{\nu+1}(x)}}} = \frac{1}{\frac{2\nu}{x} - \frac{1}{\frac{2(\nu+1)}{x} - \frac{J_{\nu+2}(x)}{J_\nu(x)}}} \\ &= \cfrac{1}{\frac{2\nu}{x} - \cfrac{1}{\frac{2(\nu+1)}{x} - \cfrac{1}{\frac{2(\nu+2)}{x} - \cfrac{1}{\ddots \; \cfrac{}{\frac{2(\nu+n)}{x} - \frac{J_{\nu+n+1}(x)}{J_{\nu+n}}}}}}},\end{aligned} \tag{36.10}$$

formally as a function. Now view this as generating numerators P_m for $1 \leq m \leq n$ and denominators Q_n for $1 \leq m \leq n$ and P^*_{n+1} and Q^*_{n+1} as the numerator and denominator when the fraction ends.

Exercise: It can be shown using the continued fraction recurrence for Q_n and repeated use of (36.7) that

$$P^*_{n+1} = J_\nu(x)$$

$$Q^*_{n+1} = J_{\nu-1}(x).$$

Remark: Observe that in

$$R = a_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{\ddots \; \cfrac{}{b_{k-1} + \cfrac{a_k}{b_k + \cfrac{a_{k+1}}{\ddots}}}}}}$$

if we set

$$b_k \mapsto \gamma b_k, \qquad a_k \mapsto \gamma a_k,$$

to get a new fraction

$$R^* = a_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{\ddots \cfrac{}{b_{k-1} + \cfrac{\gamma a_k}{\gamma b_k + \cfrac{\gamma a_{k+1}}{b_{k+1} + \ddots}}}}}}$$

and we call P_n and Q_n the numerator and denominator of R, and P_n^* and Q_n^* the numerator and denominator of R^*, then

$$P_n^* = \gamma P_n \qquad Q_n^* = \gamma Q_n,$$

for $n \geq j$.

Instead of (36.10), if we write

$$\frac{J_\nu}{J_{\nu-1}} = \cfrac{1}{\cfrac{2\nu}{x} - \cfrac{1}{\cfrac{2(\nu+1)}{x} - \cfrac{1}{\ddots \cfrac{}{\cfrac{2(\nu+n)}{x} - \cfrac{1}{(J_{\nu+n}/J_{\nu+n+1})}}}}}$$

then the new last denominator is

$$J_{\nu-1}(x)/J_{\nu+n+1}(x),$$

and the new last numerator is

$$J_\nu(x)/J_{\nu+n+1}(x).$$

If we calculate, we get that

$$\frac{J_\nu(x)}{J_{\nu-1}(x)} - \frac{P_n}{Q_n} = \frac{P_{n+1}^*}{Q_{n+1}^*} - \frac{P_n}{Q_n} = \frac{(-1)^{n+1} a_1 \cdots a_{n+1}}{Q_n Q_{n+1}^*} = \frac{J_{\nu+n+1}(x)}{Q_n J_{\nu-1}(x)}.$$

This goes to 0 as $n \to \infty$ because

(1) We have chosen $J_{\nu-1}(x) \neq 0$ (fixed).

(2) $J_{\nu+n+1}(x) \to 0$ rapidly.

(3) For large values of n, $c_n = \frac{2(v+n)}{x}$ satisfies $|c_n| \geq 2$.

Then, $Q_n \to \infty$ eventually. Thus, the iteration converges to $J_\nu(x)/J_{\nu-1}(x)$.

Theorem 36: If $J_\nu(x), J_{\nu-1}(x) \neq 0$, then

$$\frac{J_\nu(x)}{J_{\nu-1}(x)} = \cfrac{1}{\frac{2\nu}{x} - \cfrac{1}{\frac{2(\nu+1)}{x} - \cfrac{1}{\frac{2(\nu+2)}{x} - \ddots}}}$$

$$= \cfrac{x}{2\nu - \cfrac{x^2}{2(\nu+1) - \cfrac{x^2}{2(\nu+2) - \cfrac{x^2}{2(\nu+3) - \ddots}}}}$$

$$= \cfrac{x}{1 - \cfrac{x^2}{3 - \cfrac{x^2}{5 - \cfrac{x^2}{7 - \ddots}}}}.$$

6.6 Some Theorems

Theorem 37: Let $a_n, b_n \in \mathbb{Z}$, and let

$$\alpha = \frac{a_1}{b_1+}\frac{a_2}{b_2+}\cdots$$

be convergent. Suppose

$$0 < |a_n| < |b_n|, \text{ for all } n \geq N \tag{37.1}$$

and

$$|a_n| \neq |b_n| - 1, \text{ for infinitely many } n. \tag{37.2}$$

Then, α is irrational.

Proof: Define α_k, the tail of α, by

$$\alpha_K = \frac{a_k}{b_k+}\frac{a_{k+1}}{b_{k+1}+}\cdots. \tag{37.3}$$

We claim that

$$|\alpha_k| \leq 1, \text{ for all } k \geq N. \tag{37.4}$$

To realize this, note the following inequality:

$$\left|\frac{a_{k+1}}{b_{k+1}}\right| < 1,$$

and so

$$b_k - 1 < b_k + \frac{a_{k+1}}{b_{k+1}} < b_k + 1, \text{ for all } k \geq N. \tag{37.5}$$

Case 1: $(b_k > 0)$

Now,

$$b_k + \frac{a_{k+1}}{b_{k+1}} > |b_k| - 1 \geq |a_k|,$$

since $a_k, b_k \in \mathbb{Z}$.

Case 2: $(b_k < 0)$

In this case,

$$b_k + \frac{a_{k+1}}{b_{k+1}} < -|b_k| + 1 \leq -|a_k|.$$

Thus,

$$|a_k| < b_k + \frac{a_{k+1}}{b_{k+1}}.$$

So, in both cases,

$$\left|\frac{a_k}{b_k + \dfrac{a_{k+1}}{b_{k+1}}}\right| < 1. \tag{37.6}$$

Thus, the condition $\frac{a_{k+1}}{b_{k+1}} < 1$ impies (37.6). So, by iteration,

$$\left|\frac{a_k}{b_k+}\frac{a_{k+1}}{b_{k+1}+}\cdots\frac{a_{k+\ell}}{b_{k+\ell}+}\right| < 1. \tag{37.7}$$

So, letting $\ell \to \infty$, we get

$$|\alpha_k| \leq 1, \tag{37.8}$$

which proves (37.4). Next we claim that

$$|\alpha_k| < 1, \text{ infinitely often.} \tag{37.9}$$

Suppose not. Then,

$$|\alpha_k| = |\alpha_{k+1}| = 1, \text{ for all } k \geq N_1,$$

and so

$$|\alpha_k| = \left| \frac{a_k}{b_k + b_{k+1}} \right| = 1, \text{ for all } k \geq N_1.$$

Hence

$$|a_k| = |b_k + \alpha_{k+1}| \geq |b_k| - 1, \text{ for all } k \leq N_1. \tag{37.10}$$

But then, we also know from (37.1) that

$$|a_k| \leq |b_k| - 1, \text{ for all } n \geq N,$$

thus

$$|a_k| = |b_k| - 1, \text{ for all } k \geq \max(N, N_1).$$

This contradicts (37.2). So, (37.9) holds. Suppose $\alpha \in \mathbb{Q}$. We know the relation

$$\alpha = \frac{P_{k-1} + \alpha_K P_K}{Q_{k-1} + \alpha_k Q_k}.$$

Therefore,

$$\alpha_k(\alpha Q_k - P_k) = P_{k-1} - \alpha Q_{k-1}.$$

If $\alpha Q_k - P_k = 0$, then $P_{k-1} - \alpha Q_{k-1} = 0$, which implies

$$\frac{P_{k-1}}{Q_{k-1}} = \frac{P_k}{Q_k},$$

which contradicts the identity

$$\frac{P_{k-1}}{Q_{k-1}} - \frac{P_k}{Q_k} = \frac{(-1)^k a_1 \cdots a_k}{Q_k Q_{k-1}}.$$

Thus, $\alpha Q_k - P_k$ and $P_{k-1} - \alpha Q_{k-1}$ are both non-zero. Hence,

$$\alpha_k = \frac{P_{k-1} - \alpha Q_{k-1}}{\alpha Q_k - P_k} \in \mathbb{Q}.$$

Write

$$\alpha_N = \frac{A_2}{A_1}, \text{ for } A_1, A_2 \in \mathbb{Z},$$

with $|A_2| \leq |A_2|$, since $|\alpha_N| \leq 1$. Now,

$$\alpha_N = \frac{\alpha_N}{b_N + \alpha_{N+1}} = \frac{A_2}{A_1},$$

and so

$$\alpha_{N+1} = \frac{A_1 a_N - A_2 b_N}{A} = \frac{A_3}{A_2},$$

with $A_2, A_3 \in \mathbb{Z}$. But, $|\alpha_{N+1}| \leq 1$, and so $|A_3| \leq |A_2|$. Thus we get an infinite sequence of integers A_n satisfying

$$|A_1| \geq |A_2| \geq |A_3| \geq \cdots. \tag{37.11}$$

This was a consequence of $|\alpha_k| \leq 1$, for all $k \geq N$. However, $|\alpha_k| < 1$ infinitely often. Hence, in (37.11), infinitely many are to inequalities are strict. But this is a contradiction since $|A_i| \in \mathbb{Z}^+$. Thus α is irrational. $\square$

Remark: We are not claiming that the truncations yield strong approximations. Sometimes they don't.

Theorem 38: Let $\nu > 0$ be rational. Let $x \in \mathbb{Q}$, and let $J_\nu(x), J_{\nu-1}(x) \neq 0$. Then,

$$\frac{J_\nu(x)}{J_{\nu-1}(x)} \text{ is irrational.}$$

Proof: Let $x = \frac{a}{b}$. Let $\nu = \frac{c}{d}$. Then,

$$\frac{J_\nu(x)}{J_{\nu-1}(x)} = \frac{\frac{a}{b}}{2\frac{c}{d} -} \frac{\frac{a}{b}^2}{2\left(\frac{c}{d}+1\right) +} \frac{\frac{a}{b}^2}{2\left(\frac{c}{d}+2\right) -} \cdots,$$

for $a, b, c, d \in \mathbb{Z}$. We can write this as

$$\begin{aligned} \frac{a}{\frac{2bc}{d} -} \frac{a^2/b}{2\left(\frac{c}{d}+1\right) -} \cdots &= \frac{a}{2bc -} \frac{a^2 d/b}{2\left(\frac{c}{d}+1\right) -} \cdots \\ &= \frac{a}{2bc -} \frac{a^2 d}{2b\left(\frac{c}{d}+1\right) -} \cdots . \end{aligned} \tag{38.1}$$

Since $\frac{c}{d} + n \to \infty$ as $n \to \infty$, we have that the new a_n, b_n in (38.1) will be integers satisfying (37.1) and (37.2). Hence **Theorem 38** holds. □

Corollary: If $x \in \mathbb{Q}$, with $\sin x \neq 0$ and $\cos c \neq 0$, then $\tan x$ is irrational.

Proof:

$$\tan x = \frac{J_{\frac{1}{2}}(x)}{J_{-\frac{1}{2}}(x)}. \ \square$$

Corollary: π is irrational.

Proof:

$$\tan \frac{\pi}{4} = 1. \ \square$$

Remark: This is Lambert's proof of the irrationality of π. (1761)

6.7 Simple Continued Fractions

Definition: Recall a regular continued fraction has $a_n = 1$ for all $n \geq 1$. A simple continued fraction is one for which $b_n \in \mathbb{Z}^+$, and $a_0 \in \mathbb{Z}$. At this point, we relabel b_n as a_n an the continued fraction is

$$a_0 + \frac{1}{a_1+}\frac{1}{a_2+}\cdots,$$

which we denote by

$$\langle a_0, a_1, a_2, \ldots\rangle,$$

if it is infinite, and

$$\langle a_0, a_1, a_2, \ldots, a_n\rangle,$$

if it is finite.

Theorem 39: With $P_0 = a_0$, $Q_0 = 1$, $P_{-1} = 1$, $Q_{-1} = 0$, we have

$$P_n = a_n P_{n-1} + P_{n-2},$$

$$Q_n = a_n Q_{n-1} + Q_{n-2}, \text{ for } n \geq 1$$

Lemma: Let $F_0 = 0$, $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$, for $n \geq 2$. Then, $Q_n \geq F_{n+1}$.

Theorem 40: Every simple continued fraction converges to a real number.

Proof: From an earlier formula and the previous **Lemma**, we have that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{Q_n Q_{n-1}} < \infty,$$

and the fraction will have value

$$\alpha = a_0 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{Q_n Q_{n-1}},$$

if the fraction is infinite.

Note that $P_n, Q_n \in \mathbb{Z}$, and $Q_n > 0$. But also:
Theorem 41:

$$P_n Q_{n-1} - P_{n-1} Q_n = (-1)^{n-1}$$

Corollary:

$$(P_n, Q_n) = 1$$

Theorem 42: For $n \geq 2$,

$$P_n Q_{n-2} - P_{n-2} Q_n = (-1)^{n-2} a_n.$$

Proof:

$$P_n Q_{n-2} - P_{n-2} Q_n = (a_n P_{n-1} + P_{n-2}) Q_{n-2} - P_{n-2}(a_n Q_{n-1} - Q_{n-2})$$
$$= a_n (P_{n-1} Q_{n-2} - P_{n-2} Q_{n-1}) = (-1)^{n-2} a_n$$

Theorem 43: If $\alpha = \langle a_0, a_1, \ldots \rangle$ is a simple continued fraction, then P_n and Q_n satisfy the following inequalities:

$$\frac{P_0}{Q_0} < \frac{P_2}{Q_2} < \frac{P_4}{Q_4} < \cdots\cdots \frac{P_5}{Q_5} < \frac{P_3}{Q_3} < \frac{P_1}{Q_1}$$

Recall: If $\langle a_0, a_1, a_2, \ldots \rangle$ is a simple continued fraction with P_n/Q_n as the n^{th} convergent, then:

Theorem 41:

$$P_n Q_{n-1} - P_{n-1} Q_n = (-1)^{n-1}$$

Corollary:

$$\gcd(P_n, Q_n) = 1$$

Theorem 42: For $n \geq 2$,

$$P_n Q_{n-2} - P_{n-2} Q_n = (-1)^{n-2} a_n.$$

Proof:

$$P_n Q_{n-2} - P_{n-2} Q_n = (a_n P_{n-1} + P_{n-2}) Q_{n-2} - P_{n-2}(a_n Q_{n-1} - Q_{n-2}) \quad = a_n (P_{n-1} Q_{n-2} - P_{n-2} Q_{n-1}) = (-1)^{n-2} a_n.$$

Theorem 43: If $\alpha = \langle a_0, a_1, \ldots \rangle$ is a simple continued fraction, then P_n and Q_n satisfy the following inequalities:

$$\frac{P_0}{Q_0} < \frac{P_2}{Q_2} < \frac{P_4}{Q_4} < \cdots\cdots \frac{P_5}{Q_5} < \frac{P_3}{Q_3} < \frac{P_1}{Q_1}$$

If the continued fraction is infinite, we know that it converges to some value α, and by **Theorem 43**:

$$\alpha = \lim_{n \to \infty} \frac{P_{2n}}{Q_{2n}} = \lim_{n \to \infty} \frac{P_{2n-1}}{Q_{2n-1}}.$$

In particular, we have

Theorem 44: If $\langle a_0, a_1, \ldots \rangle = \alpha$ is an infinite continued fraction, then

$$0 \neq \left| \alpha - \frac{P_n}{Q_n} \right| < \frac{1}{Q_n^2}. \tag{44.1}$$

Proof: Since P_{2n}/Q_{2n} is strictly increasing and P_{2n-1}/Q_{2n-1} is strictly decreasing, we have that

$$\left| \alpha - \frac{P_n}{Q_n} \right| \neq 0.$$

Also, from **Theorem 43**, we have that

$$\left| \alpha - \frac{P_n}{Q_n} \right| < \left| \frac{P_n}{Q_n} - \frac{P_{n+1}}{Q_{n+1}} \right| = \frac{1}{Q_n Q_{n+1}} < \frac{1}{Q_n^2},$$

because $Q_{n+1} > Q_n$. □

Theorem 45: Every infinite continued fraction converges to an irrational number. Clearly, every finite continued fraction represents a rational number.

Proof: By the arguments in **Theorem 44**, P_n/Q_n are strong approximations in the Dirichlet sense to α, and so α is irrational. □

Theorem 46: Every irrational number has a unique simple continued fraction.

Proof:
Uniqueness: Let

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}. \tag{46.1}$$

be a simple continued fraction expansion of α. We will show that a_i is uniquely determined. Define

$$\alpha_n := \langle a_n, a_{n+1}, \ldots \rangle. \tag{46.2}$$

So, $\alpha = \alpha_0$. Since each α_n is itself a simple continued fraction, each α_n is irrational. Note that

$$\alpha_n > 1 \text{ for } n \geq 1. \tag{46.3}$$

Also, $\alpha = \langle a_0, a_1, \ldots, a_{n-1}, \alpha_n \rangle$. Note that

$$\alpha = a_0 + \frac{1}{\alpha_1}, \text{ with } 0 < \frac{1}{\alpha_1} < 1, \quad a_0 \in \mathbb{Z}.$$

Hence $a_0 = [\alpha]$ and $\{\alpha\} = \frac{1}{\alpha_1}$, and thus $\alpha_1 = \frac{1}{\{\alpha\}}$. So, a_0 is uniquely determined, and so is α_1. Next,

$$\alpha_1 = a_1 + \frac{1}{\alpha_2}, \text{ with } 0 < \frac{1}{\alpha_2} < 1,$$

and hence $a_1 = [\alpha_1]$. So, by iteration, the a_n are uniquely given by:

$$a_n = [\alpha_n], \qquad \alpha_{n+1} = \frac{1}{\{\alpha_n\}}. \tag{46.4}$$

Existence: Let α be irrational. Write

$$\alpha = [\alpha] + \{\alpha\} = a_0 + \{\alpha\}, \text{ with } 0 < \{\alpha\} < 1. \tag{46.5}$$

Next rewrite as:

$$\alpha = a_0 + \frac{1}{1/\{\alpha\}} = a_0 + \frac{1}{\alpha_1},$$

with $\alpha_1 = 1/\{\alpha\}$, and this yields

$$\alpha = a_0 + \frac{1}{[\alpha_1] + \{\alpha_1\}},$$

with $0 < \{\alpha_1\} < 1$. Now,

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{1/\{\alpha_1\}}}.$$

So, we may iterate this and define a_n and α_n by (46.4), at each stage obtaining an irrational α_n. Hence $\{\alpha_n\} \neq 0$. We now have the following relations

$$\alpha = \langle a_0, a_1, \ldots, a_n, \alpha_{n+1} \rangle.$$

Hence $P_m/Q_m = \langle a_0, \ldots, a_m \rangle$ for $1 \leq m \leq n$, and $P'_n/Q'_n = \langle a_0, a_1, \ldots, a_n, \alpha_{n+1} \rangle$, can be obtained from the continued fraction. Thus

$$\left.\begin{array}{rcl} P_n &=& a_n P_{n-1} + P_{n-2} \\ Q_n &=& a_n Q_{n-1} + Q_{n-2} \\ P'_n &=& \alpha_{n+1} P_n + P_{n-1} \\ Q'_n &=& \alpha_{n+1} Q_n + Q_{n-1} \end{array}\right\} \tag{46.6}$$

with $\alpha_{n+1} > 1$. Thus,

$$\begin{aligned}\alpha - \frac{P_n}{Q_n} = \frac{P'_n}{Q'_n} - \frac{P_n}{Q_n} &= \frac{\alpha_{n+1}P_n + P_{n-1}}{\alpha_{n+1}Q_n + Q_{n-1}} - \frac{P_n}{Q_n} \\ &= \frac{P_{n-1}Q_n - P_nQ_{n-1}}{Q_n(\alpha_{n+1}Q_n + Q_{n-1})} \\ &= \frac{(-1)^n}{Q_n(\alpha_{n+1}Q_n + Q_{n-1})} \xrightarrow{n\to\infty} 0.\end{aligned}$$

Thus the continued fraction converges to α. $\square$

Remark: The existence proof has given an exact measure of the difference $\alpha - P_n/Q_n$. If a_n are large, then the approximations improve.

6.8 Finite Continued Fraction Expansion of Rationals

If we apply the above contradiction as in the existence, we cannot get an infinite expansion, because it would force α to be irrational. Thus,

$$\alpha = \langle a_0, a_1, \ldots, a_n \rangle,$$

which is a finite expansion which stops only because $\{a_n\} = 0$, i.e., $\alpha_n \in \mathbb{Z}$ so $[\alpha_n] = a_n$.

If $a_n > 1$, we could write a_n as $(a_n - 1) + \frac{1}{1}$, which gives us another continued fraction one longer. This does not come out of our earlier algorithm. So, there are exactly two continued fraction expansions for each rational number, (which we restate below as a theorem), but only one comes out of our algorithm.

Theorem 47: Every rational number has exactly two continued fraction expansions: one ending in $a_n \geq 2$, and one ending in $a_{n+1} = 1$. If we add the condition that the finite continued fraction expansion does not end in a 1, then we have that finite continued fraction expansions are unique. (This is similar to the restriction that a decimal does not end with only 9s.)

Question: Can we classify reals (irrationals) by their continued fraction expansions?

We will prove α is represented as a periodic continued fraction if and only if α is a quadratic irrational.

6.9 Minkowiski Question Mark Function

Pick $\alpha \in \mathbb{Q}$, and write $\alpha = a_0.b_0b_1 \cdots b_m\overline{b_{m+1} \cdots b_{m+s}}$. Now define

$$\beta = \beta(\alpha) = a_0 + \cfrac{1}{b_0 + \cfrac{1}{b_1 + \cfrac{1}{\ddots}}}.$$

Then β is a quadratic irrational. Minkowski's Question Mark function used this construction to map quadratic irrationals to rationals. This is an example of a singular function.

6.10 Continued Fraction for e

Theorem 48: (Euler)

$$e = \langle 2, 1, 2, 1, 1, 4, 1, 1, 6, 1, \ldots \rangle, \qquad \text{i.e.,} \qquad \left| e - \frac{P_n}{Q_n} \right| < \frac{1}{\kappa Q_n^2},$$

with κ as large as possible. We will prove this by showing that

$$\frac{e+1}{e-1} = \langle 2, 6, 10, 14, \ldots \rangle,$$

which is much easier. Then we can relate the two fractions.

Proof: First we prove that

$$\frac{e+1}{e-1} = \langle 2, 6, 10, 14, \ldots \rangle.$$

Recall that

$$\tan(x) = \cfrac{x}{1 - \cfrac{x^2}{3 - \cfrac{x^2}{5 - \ddots}}},$$

and so,

$$\tan(ix) = \cfrac{ix}{1 + \cfrac{x^2}{3 + \cfrac{x^2}{5 + \ddots}}}. \tag{48.1}$$

Since $\tan(ix) = i \tanh(x)$, we get that

$$\tanh(x) = \cfrac{x}{1 + \cfrac{x^2}{3 + \cfrac{x^2}{5 + \ddots}}}. \tag{48.2}$$

Note that

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}. \tag{48.3}$$

Since $\tanh(x)$ is irrational for $0 \neq x \in \mathbb{Q}$, we conclude from (48.3) that $(e^x)^2$ is irrational for rational $x \neq 0$. Put $x = \frac{1}{2}$ in (48.2) to get

$$\frac{e-1}{e+1} = \cfrac{1/2}{1 + \cfrac{1/4}{3 + \cfrac{1/4}{5 + \ddots}}} = \cfrac{1}{2 + \cfrac{1/2}{3 + \cfrac{1/4}{5 + \ddots}}} = \cfrac{1}{2 + \cfrac{1}{6 + \cfrac{1}{10 + \ddots}}}. \tag{48.4}$$

So,

$$\frac{e+1}{e-1} = \langle 2, 6, 10, \cdots, 2(2n+1), \cdots \rangle. \tag{48.5}$$

Let P_n/Q_n denote the n^{th} convergent in the proposed continued fraction for e and let P'_n/Q'_n denote the n^{th} convergent in (48.5). We now claim the following:

Lemma: For each $k \in \mathbb{Z}^+$, we have that

$$P_{3k+1} = P'_k + Q'_k, \qquad \text{and} \qquad Q_{3k+1} = P'_k - Q'_k. \tag{48.6}$$

Proof: We will establish the first equality in (48.6). The second is similar, and it should be completed by the reader as an exercise.

Set $S_k = P_{3k+1}$ and $T_k = P'_k + Q'_k$. We will show by induction that $S_k = T_k$.

Initial Step:

$S_0 = P_1 =$ the numerator of $2 + \frac{1}{1} = 3$.
$T_0 = P'_0 + Q'_0 = 2 + 1 = 3$.
Hence $S_0 = T_0$.

Similarly, $S_1 = P_4 =$ the numerator of $2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1}}}} = \frac{19}{7}$, thus $S_1 = 19$.

Now, $T_1 = P'_1 + Q'_1 = 13 + 6 = 19$.
Hence $S_1 = T_1$.

Inductive Step:

From the fraction (48.5), we see that

$$P'_k = 2(2k+1)P'_{k-1} + P'_{k-2},$$
$$Q'_k = 2(2k+1)Q'_{k-1} + Q'_{k-2}.$$

Therefore,

$$T_k = 2(2k+1)T_{k-1} + T_{k-2}. \tag{48.7}$$

Observe that

$$\left.\begin{array}{rcrcl} P_{3k-3} &=& P_{3k-4} &+& P_{3k-5}, \\ P_{3k-2} &=& P_{3k-3} &+& P_{3k-4}, \\ P_{3k-1} &=& 2kP_{3k-2} &+& P_{3k-3}, \\ P_{3k} &=& P_{3k-1} &+& P_{3k-2}, \\ P_{3k+1} &=& P_{3k} &+& P_{3k-1}. \end{array}\right\}. \tag{48.8}$$

Notice also that

$$P_{3k-3} - P_{3k-2} + 2P_{3k-1} + P_{3k} + P_{3k+1} = (P_{3k-4} + P_{3k-5}) - (P_{3k-3} + P_{3k-4}) + (4kP_{3k-2} + 2P_{3k-3}) + (P_{3k-1} + P_{3k-2}) + (P_{3k} + P_{3k-1}) \tag{48.9}$$

By cancellation,

$$P_{3k+1} = (4k+2)P_{3k-2} + P_{3k-5},$$

i.e. $S_k = 2(2k+1)S_{k-1} + S_{k-2}$. Thus comparing (48.7) and (48.9), we see that T_k and S_k satisfy the same recurrences. We checked at $k = 0, 1$. Thus $T_k = S_k$, as claimed. □

From the **Lemma**, we know that

$$\lim_{k\to\infty} \frac{P_k}{Q_k} = \lim_{k\to\infty} \frac{P_{3k+1}}{Q_{3k+1}} = \lim_{k\to\infty} \frac{P_k + Q_k}{P'_k - Q'_k}, \text{ and so}$$

$$\lim_{k\to\infty} \frac{P_k}{Q_k} = \lim_{k\to\infty} \frac{P_k + Q_k}{P'_k - Q'_k} = \lim_{k\to\infty} \frac{\frac{P'_k}{Q'_k} + 1}{\frac{P'_k}{Q'_k} - 1} = \frac{\frac{e+1}{e-1} + 1}{\frac{e+1}{e-1} - 1} = e. \ \square$$

6.11 Irrational Numbers and Continued Fraction Convergence

Lemma: Let $\frac{a}{b}$ and $\frac{c}{d}$ be two rationals satisfying

$$|ad - bc| = 1. \tag{49.1}$$

Let θ be an irrational that lies between $\frac{a}{b}$ and $\frac{c}{d}$. Let $\frac{a}{b}$ be the fraction with the smaller denominator. Then,

$$theta - \frac{a}{b} < \frac{1}{b^2}. \tag{49.2}$$

Proof: Note that

$$\left|\theta - \frac{a}{b}\right| < \left|\frac{a}{b} - \frac{c}{d}\right| = \frac{|ad - bc|}{bd} = \frac{1}{bd} < \frac{1}{b^2}. \ \square$$

Remark: Let θ be irrational, and let p_n/q_n be a convergent to θ. Consider the next convergent p_{n+1}/q_{n+1}. We know that θ lies between the two, and we also know that

$$|p_n q_{n+1} - p_{n+1} q_{n+1}| = 1. \tag{49.3}$$

In addition, $q_n < q_{n+1}$. Thus, by the lemma,

$$\left|\theta - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2} \tag{49.4}$$

is valid for every convergent.

Lemma: If $\frac{a}{b}$ and $\frac{c}{d}$ are two rationals satisfying (49.1), and if θ is an irrational that lies between them, then one among $\frac{a}{b}$ and $\frac{c}{d}$ (call it $\frac{p}{q}$), would satisfy

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{2q^2}. \tag{49.5}$$

Proof: Suppose

$$\left.\begin{aligned} \left|\theta - \frac{a}{b}\right| &> \frac{1}{2b^2}, \\ \left|\theta - \frac{c}{d}\right| &> \frac{1}{2d^2} \end{aligned}\right\} \tag{49.6}$$

Then

$$\left|\frac{a}{b} - \frac{c}{d}\right| = \left|\theta - \frac{a}{b}\right| + \left|\theta - \frac{c}{d}\right| > \frac{1}{2b^2} + \frac{1}{2d^2} \tag{49.7}$$

which is equivalent to

$$2bd > d^2 + b^2,$$

which is true if and only if

$$0 > (b - d)^2,$$

which is a contradiction. Hence the lemma. $\square$

Theorem 49: For an irrational θ, one out of every two consecutive convergents will satisfy:

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{2q^2}.$$

Note that we are not specifying which among a pair of consecutive convergents will satisfy (49.5). It could be the one with the larger denominator, and then both consecutive convergents satisfy (49.2). This theorem is of interest because it has a partial converse.

Theorem 50: If θ is irrational, and $\frac{p}{q}$ is rational, such that

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{2q^2},$$

then $\frac{p}{q}$ is a convergent to θ.

Proof: First write

$$\left|\theta - \frac{p}{q}\right| = \frac{\epsilon\eta}{q^2}, \tag{50.1}$$

with $\epsilon = \pm 1$ and $0 < \eta < 1/2$. Next, write

$$\frac{p}{q} = \langle a_0, a_1, \ldots, a_n \rangle,$$

and choose n such that $\epsilon = (-1)^n$, i.e.,

$$\operatorname{sgn} \frac{p}{q} = (-1)^n.$$

(Here we are choosing between the two ways to end a rational continued fraction to satisfy the above condition.)

Now, we can rewrite (50.1) as

$$\theta - \frac{P_n}{Q_n} = \frac{(-1)^n \eta}{Q_n^2}. \tag{50.2}$$

Let $P_{n-1}/Q_{n-1} = \langle a_0, a_1, \ldots, a_{n-1} \rangle$. Then

$$|P_n Q_{n-1} - P_{n-1} Q_n| = 1.$$

Thus, there exists ω such that

$$\theta = \frac{P_n \omega + P_{n-1}}{Q_n \omega + Q_{n-1}}. \tag{50.3}$$

A nice way to see (50.3) is to write it as

$$\begin{bmatrix} \theta \\ 1 \end{bmatrix} = \begin{bmatrix} P_n & P_{n-1} \\ Q_n & Q_{n-1} \end{bmatrix} \begin{bmatrix} \omega \\ 1 \end{bmatrix}, \tag{50.4}$$

which is the same as

$$\begin{bmatrix} \omega \\ 1 \end{bmatrix} = \begin{bmatrix} P_n & P_{n-1} \\ Q_n & Q_{n-1} \end{bmatrix} \begin{bmatrix} \theta \\ 1 \end{bmatrix}. \tag{50.5}$$

So, from (50.2) and (50.3), we get that

$$\begin{aligned} \frac{(-1)^n \eta}{Q_n^2} &= \theta - \frac{P_n}{Q_n} \\ &= \frac{P_n \omega + P_{n-1}}{Q_n \omega + Q_{n-1}} - \frac{P_n}{Q_n} \\ &= \frac{P_{n-1} Q_n - P_n Q_{n-1}}{(Q_n \omega + Q_{n-1}) Q_n} \\ &= \frac{(-1)^n}{(Q_n \omega + Q_{n-1}) Q_n}. \end{aligned} \tag{50.6}$$

This yields

$$\eta = \frac{Q_n}{Q_n\omega + Q_{n-1}}, \tag{50.7a}$$

and hence

$$\omega = \frac{Q_n - \eta Q_{n-1}}{\eta Q_n}. \tag{50.7b}$$

Since $0 < \eta < 1/2$ and $0 < Q_{n-1} < Q_n$, and by (50.3), we have that

$$\omega > 1 \text{ and } \omega \text{ is irrational.} \tag{50.8}$$

Now, write ω as a continued fraction

$$\omega = \langle a_{n+1}, a_{n+2}, \cdots, \rangle. \tag{50.9}$$

From (50.3), we see that $\theta = \langle a_0, a_1, \cdots, a_n, \omega \rangle$. But, we see that this is equal to $\theta = \langle a_0, a_1, \cdots, a_n, a_{n+1}, \cdots, \rangle$ as a simple continued fraction. Since such an expansion is unique, it is the simple continued fraction expansion of θ, and P_n/Q_n is the n^{th} convergent. $\square$

Chapter 7

Best Approximations

7.1 Definition and Some Theorems

Remark: If $\alpha = \langle a_0, a_1, \cdots \rangle$ and $\alpha_n = \langle a_n, a_{n+1}, \cdots \rangle$, then we have seen that

$$\alpha = \frac{\alpha_n P_{n-1} + P_{n-2}}{\alpha_n Q_{n-1} + Q_{n-2}}, \tag{51.1}$$

and,

$$a_n = [\alpha_n]. \tag{51.2}$$

From (51.1) we get

$$\alpha_n = \frac{P_{n-2} - \alpha Q_{n-2}}{\alpha Q_{n-1} - P_{n-1}}. \tag{51.3}$$

So, we arrive at:

Corollary: If $\alpha = \langle a_0, a_1, \cdots \rangle$, then the partial quotients a_n are given by

$$a_n = \left[\frac{|\alpha Q_{n-2} - P_{n-2}|}{|\alpha Q_{n-1} - P_{n-1}|} \right]. \tag{51.4}$$

We begin by defining the norm of a real number x by

$$\|x\| = \min_{n \in \mathbb{Z}} |x - n|, \tag{51.5}$$

so that $\|x\|$ is the distance between x and the integer nearest to x. This norm satisfies the triangle inequality

$$\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|,$$

and $\|x\| = 0$ is and only if $x \in \mathbb{Z}$. Hence, this norm is a metric on the circle group.

Since we wish to deal with best approximations to both rationals and irrationals, we will reformulate Dirichlet's theorem (**Theorem 10**) suitably.

Theorem 51: Let $\theta, Q \in \mathbb{R}$ with $Q > 1$. Then, there exists an integer q such that

$$0 < q < Q \text{ and } \|q\theta\| \leq \frac{1}{Q}.$$

Remark: Theorem 51 is best-possible by choosing $\theta = \frac{1}{Q}$, for $Q \in \mathbb{Z}^+$.

Proof: First assume that $Q \in \mathbb{Z}$. Consider the $Q+1$ numbers

$$\{0,1\} \cup \{q\theta \mid 0 < q < Q,\ q \in \mathbb{Z}\} \subset [0,1]. \tag{51.8}$$

Partition [0,1] as

$$[0,1] = \left[0, \frac{1}{Q}\right) \cup \left[\frac{1}{Q}, \frac{2}{Q}\right) \cup \cdots \cup \left[\frac{Q-1}{Q}, 1\right]. \tag{51.9}$$

By the pigeonhole principle, two members from the list in (51.8) lie in the same subset in (51.9), i.e., there exist integers r_1, s_1, r_2, s_2 such that

$$|(r_1\theta - s_1) - (r_2\theta - s_2)| \leq \frac{1}{Q},$$

with $0 \leq r_1 \neq r_2 < Q$. So,

$$q = |r_1 - r_2|$$

does the job.

Now, if Q is not an integer, then apply the above argument for $[Q]+1$ and note that $q < [Q]+1$, and therefore $q \leq [Q] < Q$. In fact, we get that $\|q\theta\| \leq \frac{1}{[Q]+1} < \frac{1}{Q}$. □

Definition: If $\theta \in \mathbb{R}$, a rational p/q is called a best approximation of θ if:

$$\|q\theta\| < \|q'\theta\|, \text{ for all } 0 < q' < q. \tag{52.1}$$

Remark: Clearly, $q = 1$ yields a best approximation with $p =$ the nearest integer to θ. Define $q_1 = 1$ and note that $\|q_1\theta\| \leq 1/2$. Also, $\|q_1\theta\| = 0$ if and only if $\theta \in \mathbb{Z}$, in which case our process stops. If $\|q_1\theta\| \neq 0$, then choose $Q > \dfrac{1}{\|q_1\theta\|}$, so that

$$\frac{1}{Q} < \|q_1\theta\|. \tag{52.2}$$

By, **Theorem 51**, the set of Q for which

$$\|q\theta\| < \frac{1}{Q} < \|q_1\theta\|, \text{ for all } 0 < q < Q \tag{52.3}$$

is non-empty. From among these q satisfying (52.3), choose q_2 to be minimal. Hence,

$$\|q_2\theta\| = |q_2\theta - p_2| < |q_1\theta| \leq \frac{1}{2}.$$

If $\|q_2\theta\| = 0$, we stop here. Otherwise, we use (52.2) to iterate. This process generates the sequence of best approximations.

Recall: Given $\theta \in \mathbb{R}$, we generated a sequence of positive integers q_n starting with $q_1 = 1$, that satisfied the following:

$$1 = q_1 < q_2 < q_3 < \cdots \tag{52.4}$$

and their companions

$$p_1, p_2, \cdots, p_n \in \mathbb{Z},$$

such that

$$\left.\begin{array}{c} \|q_n\theta\| = |q_n\theta - p_n|, \\ \|q_{n+1}\theta\| < \|q_n\theta\|, \\ \|q_n\theta\| < \|q\theta\|, \text{ for all } 1 \leq q < q_{n+1}. \end{array}\right\} \tag{54.5}$$

We call each p_n/q_n as a best approximation to θ. Note that the sequence q_n can be finite and would end at n when $\|q_n\theta\| = 0$.

Theorem 52: For each n for which q_{n+1} is defined $q_{n+1}\theta - p_{n+1}$ and $q_n\theta - p_n$ are of opposite sign.

Proof: Suppose $q_{n+1}\theta - p_{n+1}$ and $q_n\theta - p_n$ have the same sign. We know that $|q_{n+1}\theta - p_{n+1}| < |q_n\theta - p_n|$, by the definition of best approximation. Since, they have the same sign, $(q_{n+1} - q_n)\theta - (p_{n+1} - p_n) = q'\theta - p'$, with $q' = q_{n+1} - q_n$ and $p' = p_{n+1} - p_n$. Now, this satisfies

$$|q'\theta - p'| < |q_n\theta - p_n|, \tag{52.6}$$

with $q' < q_{n+1}$, if $q_{n+1}\theta - p_{n+1} \neq 0$. This violates (52.5), and hence the theorem holds when $q_{n+1}\theta - p_{n+1}\theta \neq 0$.

If $q_{n+1}\theta - p_{n+1} = 0$, the theorem is trivially true. □

Remark: We can rewrite Theorem 52 as:

If $q_{n+1}\theta - p_{n+1} \neq 0$, then

$$(q_n\theta - p_n)(q_{n+1}\theta - p_{n+1}) < 0. \tag{52.7}$$

From the minimality of q_{n+1} and the choice $Q = q_{n+1}$ in **Theorem 51**, we get that

$$\|q_n\theta\| < \frac{1}{Q} = \frac{1}{q_{n+1}}. \tag{53.1}$$

Hence,

$$\left.\begin{aligned} q_n\|q_n\theta\| < q_{n+1}\|q_n\theta\| < 1, \quad \text{and} \\ \left|\theta - \frac{p_n}{q_n}\right| < \frac{1}{q_nq_{n+1}}. \end{aligned}\right\} \tag{53.2}$$

Theorem 53: If θ is real and p_n/q_n the sequence of best approximations, then

(i) If θ is rational, then $\theta = p_N/q_N$ for some N, i.e. $\|q_N\theta\| = 0$.

(ii) If θ is irrational, the sequence is infinite and yields strong approximations in the Dirichlet sense.

Lemma: If p_n/q_n and p_{n+1}/q_{n+1} are successive best approximations to a real θ, then

$$|p_nq_{n+1} - p_{n+1}q_n| = 1.$$

Proof: We begin by noting that

$$q_{n+1}p_n - q_np_{n+1} = q_n(q_{n+1}\theta - p_{n+1}) = q_{n+1}(q_n\theta - p_n) \tag{54.1}$$

We know that $|q_{n+1}\theta - p_{n+1}| < |q_n\theta - p_n|$, and so

$$|q_n(q_{n+1}\theta - p_{n+1})| < |q_{n+1}(q_n\theta - p_n)|. \tag{54.2}$$

Therefore,

$$\begin{aligned} &0 < |p_nq_{n+1} - p_{n+1}q_n| < q_n\|q_{n+1}\theta\| + q_{n+1}\|q_n\theta\|, \\ &0 < |p_nq_{n+1} - p_{n+1}q_n| < 2q_{n+1}\|q_n\theta\|, \\ &0 < |p_nq_{n+1} - p_{n+1}q_n| < 2. \end{aligned}$$

by (53.2). But, $|p_nq_{n+1} - p_{n+1}q_n| \in \mathbb{Z}^+$. Hence $|p_nq_{n+1} - p_{n+1}q_n| = 1$. □

Theorem 54: Let p_n/q_n and p_{n+1}/q_{n+1} be successive best approximations to a real number θ. Then,

(i) $q_{n+1}p_n - q_np_{n+1}$ and $q_n\theta - p_n$ have opposite sign.

(ii) $q_{n+1}p_n - q_np_{n+1} = -(q_np_{n-1} - q_{n-1}p_n)$

(iii) $q_n\|q_{n+1}\theta\| + q_{n+1}\|q_n\theta\| = 1$.

Proof: By (51.4) and (54.2), we see that

$$\operatorname{sgn} q_{n+1}p_n - q_np_{n+1} = \operatorname{sgn} -q_{n+1}(q_n\theta - p_n) = -\operatorname{sgn} q_n\theta - p_n.$$

Hence (i) is true. (ii) follows directly from (i) and (52.7). To see (iii), use (52.7) again and rewrite (54.1) as

$$|q_{n+1}p_n - q_np_{n+1}| = q_n|q_{n+1}\theta - p_{n+1}| + q_{n+1}|q_n\theta - p_n| = q_n\|q_{n+1}\theta\| + q_{n+1}\|q_n\theta\|,$$

and use the above **Lemma**. □

Theorem 55: If p_n/q_n is the sequence of best approximations to θ, then there exists $a_n \in \mathbb{Z}^+$ such that

$$p_{n+1} = a_np_n + p_{n_1}, \tag{55.1}$$

$$q_{n+1} = a_nq_n + q_{n-1}, \tag{55.2}$$

$$|q_{n-1}\theta - p_{n-1}| = a_n|q_n\theta - p_n| + |q_{n+1}\theta - p_n|. \tag{55.3}$$

Proof: From **Theorem 54**, we have that

$$(q_{n+1} - q_{n-1})p_n = (p_{n+1} - p_{n-1})q_n. \tag{55.4}$$

By the **Lemma**, we have that $\gcd(p_n, q_n) = 1$. Therefore, with $p_n \mid (p_{n+1} - p_{n-1})q_n$, we have that

$$p_n \mid (p_{n+1} - p_{n-1}). \tag{55.5}$$

Thus, there exists and integer a_n such that

$$p_{n+1} - p_{n-1} = a_np_n. \tag{55.6}$$

Similarly, $q_n \mid (q_{n+1} - q_{n-1})p_n$, so we have that there exists b_n such that

$$b_nq_n = q_{n+1} - q_{n-1}. \tag{55.7}$$

Now, by (55.4), (55.6), and (55.7),

$$b_nq_np_n = q_np_nq_n,$$

and so $a_n = b_n$. Thus the same integer satisfies both (55.1) and (55.2). Note that (55.2) together with (52.4) forces that $a_n > 0$.

Lastly, to prove (55.3), multiply (55.2) by θ and subtract (55.1). Now,

$$q_{n+1}\theta - p_{n+1} = a_n(q_n\theta - p_n) + (q_{n-1}\theta - p_{n-1}).$$

$$(q_{n+1}\theta - p_{n+1}) - a_n(q_n\theta - p_n) = q_{n_1}\theta - p_{n-1}.$$

Since

$$\operatorname{sgn} q_{n-1}\theta - p_{n-1} = \operatorname{sgn} q_{n+1}\theta - p_{n+1} = -\operatorname{sgn} q_n\theta - p_n,$$

and $a_n > 0$, (55.3) follows. □

Consequence: We can rewrite (55.3) as:

$$\frac{|q_{n-1}\theta - p_{n-1}|}{|q_n\theta - p_n|} = a_n + \frac{|q_{n+1}\theta - p_{n+1}|}{|q_n\theta - p_n|}.$$

Since $a_n \in \mathbb{Z}$ and $|q_{n+1}\theta - p_{n+1}| < |q_n\theta - p_n|$, we conclude that

$$a_n = \left[\frac{|q_{n-1}\theta - p_{n-1}|}{|q_n\theta - p_n|}\right].$$

Compare with

$$a_n = [\alpha_n] = \left[\frac{|\alpha Q_{n-2} - P_{n-2}|}{|\alpha Q_{n-1} - P_{n-1}|}\right],$$

which was **Theorem 51**, line (51.3). With an index shift by 1, we see that they are the same.

7.2 A Procedure to Generate Best Approximations

Note that if p/q is a best approximation to θ, then

$$\frac{p}{q} - [\theta]$$

is a best approximation to $\{\theta\} = \theta - [\theta]$. So, we will concentrate on $0 < \theta < 1$. This discussion will be broken into two cases: (1) $0 < \theta \leq 1/2$, and (2) $1/2 < \theta < 1$.

Case 1: $(0 < \theta < 1/2)$

With $q_1 = 1$, the nearest integer is 0. Set $p_1 = 0$. Thus,

$$q_1\theta - p_1 = \theta > 0, \quad \|\theta\| \leq \frac{1}{2}. \tag{56.1}$$

Since $|q_{n+1}p_n - p_{n+1}q_n| = 1(\bigstar)$, taking $n = 1$ gives $|q_2p_1 - p_2q_1| = 1$, and hence $|p_2| = 1$, and since $\theta > 0$ we have

$$p_2 = 1. \tag{56.2}$$

With $p_2 = 1$ and $p_1 = 0$, use $p_2 = a_1p_1 + p_0$ to get

$$p_0 = 1. \tag{56.3}$$

At this stage define $q_0 := 0$. Since $q_1 = 1$ and $q_2 = a_1q_1 + q_0$, we have that

$$q_2 = a_1. \tag{56.4}$$

So, we need to know a_1 in order to determine q_2. For this we use

$$|q_0\theta - p_0| = a_1|q_1\theta - p_1| + |q_2\theta - p_2|, \tag{56.5}$$

and therefore,

$$1 = a_1\theta + \|q_2\theta\|.$$

So,

$$\frac{1}{\theta} = a_1 + \frac{\|q_2\theta\|}{\|q_1\theta\|}. \tag{56.6}$$

Since $\|q_2\theta\| < \|q_1\theta\|$, we have that

$$a_1 = [1/\theta]. \tag{56.7}$$

So, we have determined a_1 and thus q_2. We have the formula

$$a_0 = \left[\frac{\|q_{n-1}\theta\|}{\|q_n\theta\|}\right], \text{ for } n \geq 2$$

from last class. So, since we have q_1 and q_2, we can calculate a_2, and we can start our recurrence since we now have a_1 and a_2, and the recurrences

$$q_{n+1} = a_nq_n + q_{n-1}$$

$$p_{n+1} = a_np_n + p_{n-1},$$

so we can calculate a_2, then q_3, then p_3, then a_3, and so on.

Case 2: $(1/2 < \theta < 1)$

With $q_1 = 1$, the nearest integer is 1 (not 0) and therefore $p_1 = 1$, and thus

$$q_1\theta - p_2 = \theta - 1 < 0. \tag{56.8}$$

We know that $|q_2p_1 - p_2q_1| = 1$, and therefore

$$|q_2 - p_2| = 1 \tag{56.9}$$

But, we also know that $\operatorname{sgn} q_2p_1 - p_2q_1 = -\operatorname{sgn} q_1\theta - p_1$, and the right hand term is less than zero. Thus

$$q_2 - p_2 = 1. \tag{56.10}$$

If we define

$$q_0 := 1 \text{ and } p_0 := 0, \tag{56.11}$$

then $|q_1p_0 - q_0p_1| = 1$, which fits with (★). Also, $q_1p_0 - p_1q_0 = -1$ and is of opposite sign to $q_0\theta - p_0 = 1 \cdot \theta - 0 = \theta$. In addition, $p_2 = a_1p_1 + p_0$ implies that

$$p_2 = a_1, \tag{56.12}$$

and $q_2 = a_1q_1 + q_0$ implies that

$$q_2 = a_1 + 1. \tag{56.13}$$

This is all consistent with (56.10). So, (56.11) is a valid choice that yields consistent properties between Case 1 and Case 2.

Similarly, $p_{-1} = 1$ and $q_{-1} = 0$ (56.14) are valid choices as well. With these choices, we have

$$|q_{-1}\theta - p_{-1}| = a_0|q_0\theta - p_0| + |q_1\theta - p_1|.$$

Therefore,from (★), we get

$$1 = a_0\theta + |\theta - 1|.$$

Thus,

$$\frac{1}{\theta} = a_0 + \frac{|\theta - 1|}{\theta}. \tag{56.15}$$

Since $0 < |\theta - 1| < \theta$, we have that $a_0 = [1/\theta]$. (56.16)
Hence, we're back to (56.7), and we iterate from here.

In order to fit both cases into one theorem, we shift the index in Case 2 by 1 to get:

Theorem 56: Let $0 < \theta < 1$. Define p_n, q_n, a_n by:

$$p_0 = 1, q_0 = 0,$$

$$p_1 = 0, q_1 = 1,$$

$$p_{n+1} = a_np_n + p_{n-1},$$

$$q_{n+1} = a_nq_n + q_{n-1},$$

where

$$a_n = \left[\frac{|q_{n-1}\theta - p_{n-1}|}{|q_n\theta - p_n|}\right],$$

if $q_n\theta \neq p_n$. If $q_n\theta = p_n$, then the procedure stops with a_{n-1}. Then, the p_n/q_n are (i) the best approximations to θ for $n \geq 1$ if $0 < \theta \leq 1/2$, and (ii) the best approximations to θ for $n \geq 2$, if $1/2 < \theta < 1$. Moreover $(-1)^{n+1}(q_n\theta - p_n) \geq 0$, and $q_{n+1}p_n - q_np_{n+1} = (-1)^n$.

7.3 Connection with Continued Fractions

We denoted by P_n and Q_n the n^{th} numerator and denominator of the fraction. The connection is that

$$q_n = Q_{n-1} \text{ and } p_n = P_{n-1}. \tag{57.1}$$

Theorem 57: Given an irrational θ the convergents P_n/Q_n are (i) the best approximations to θ for $n \geq 0$ if $0 < \theta \leq 1/2$, and (ii) the best approximations to θ for $n \geq 1$ if $1/2 < \theta < 1$.

Proof: To generate the continued fraction process, write $\theta = [\theta] + \{\theta\}$. If $0 < \{\theta\} \leq 1/2$, then $[\theta]$ is the nearest integer. If $1/2 < \{\theta\} < 1$, then the nearest integer is $[\theta] + 1$. In this case, note that

$$\theta = [\theta] + \{\theta\} = [\theta] + \frac{1}{1/\{\theta\}}.$$

Then, the next convergent is given by

$$[\theta] + \frac{1}{[1/\{\theta\}]} = [\theta] + 1/1 = [\theta] + 1,$$

with $[\theta] + 1$ equals the nearest integer. □

Remark: If $\theta \in \mathbb{Q}$, then θ has <u>two</u> continued fraction expansions, then

$$\theta = \langle a_0, a_1, \ldots, a_{N-1}, a_N \rangle, \text{ with } a_N \geq 2. \tag{58.1a}$$

$$\theta = \langle a_0, a_1, \ldots, a_{N-1}, a_N - 1, 1 \rangle, \text{ with } a_{N+1} = 1. \tag{58.1b}$$

Theorem 58: If $\theta = p/q$, has the expansions (58.1a) and (58.1b) above, then (i) ALL convergents to (a) are best approximations for $n \geq 1$, and (ii) ALL convergents to (b) are best approximations <u>except</u> $\langle a_0, a_1, \ldots, a_{N-1}, a_N - 1 \rangle$.

Chapter 8

Equivalence of Real Numbers

8.1 Definition

Definition: Two real numbers θ and θ' are said to be equivalent if there exists integers r, s, t, u such that

$$|ru - ts| = 1, \tag{59.1}$$

and

$$\theta = \frac{r\theta' + s}{t\theta' + u}. \tag{59.2}$$

We may write (59.2) in matrix vector form as

$$\begin{bmatrix} \theta \\ 1 \end{bmatrix} = \begin{bmatrix} r & s \\ t & u \end{bmatrix} \begin{bmatrix} \theta' \\ 1 \end{bmatrix}, \tag{59.3}$$

and utilize the group properties of

$$\mathcal{S} := \left\{ \begin{bmatrix} r & s \\ t & u \end{bmatrix} \mid r, s, t, u \in Z, \;\; |rs - tu| = 1 \right\}.$$

The relation defined by (59.2) is an equivalence relation. For reflexivity, let

$$\begin{bmatrix} r & s \\ t & u \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

To see symmetry, consider

$$\begin{pmatrix} \theta' \\ 1 \end{pmatrix} = \begin{pmatrix} r & s \\ t & u \end{pmatrix}^{-1} \begin{pmatrix} \theta \\ 1 \end{pmatrix}.$$

For transitivity, assume

$$\begin{pmatrix} \theta \\ 1 \end{pmatrix} = \begin{pmatrix} r & s \\ t & u \end{pmatrix} \begin{pmatrix} \theta' \\ 1 \end{pmatrix},$$

and

$$\begin{pmatrix} \theta' \\ 1 \end{pmatrix} = \begin{pmatrix} r' & s' \\ t' & u' \end{pmatrix} \begin{pmatrix} \theta'' \\ 1 \end{pmatrix},$$

and therefore

$$\begin{pmatrix} \theta \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} r & s \\ t & u \end{pmatrix} \begin{pmatrix} r' & s' \\ t' & u' \end{pmatrix}}_{\in S} \begin{pmatrix} \theta'' \\ 1 \end{pmatrix}. \tag{59.4}$$

8.2 Connection with Continued Fractions

If

$$\theta = \langle a_0, a_1, \cdots \rangle, \tag{59.5}$$

with

$$P_m/Q_m = \langle a_0, a_1, \cdots, a_m \rangle, \tag{59.6}$$

and

$$\theta = \langle a_0, a_1, \cdots, a_{n-1}, \theta_n \rangle, \tag{59.7}$$

then

$$\theta = \frac{\theta_n P_{n-1} + P_{n-2}}{\theta_n Q_{n-1} + Q_{n-2}}, \tag{59.8}$$

where

$$|P_{n-1}Q_{n-2} - P_{n-2}Q_{n-1}| = 1. \tag{59.9}$$

Thus,

$$\theta \sim \theta_n, \text{ for all } n. \tag{59.10}$$

This leads to:

Theorem 59: Two real numbers θ and θ' are equivalent if and only if there exists integers $a_0, \cdots, a_m$ and $b_0, \cdots, b_m$, and $c_1, c_2, \cdots$ all in $\mathbb{Z}^+$ such that

$$\theta = \langle a_0, \cdots, a_m, c_1, c_2, \cdots \rangle \tag{59.11}$$

$$\theta' = \langle b_0, \cdots, b_m, c_1, c_2, \cdots \rangle. \tag{59.12}$$

In particular, any two rationals in reduced form are equivalent.

Proof: ($\Longrightarrow$) We are given (59.11) and (59.12). Define

$$\theta'' := \langle c_1, c_2, \cdots \rangle. \tag{59.13}$$

So, by (59.10),

$$\theta \sim \theta'', \ \theta \sim \theta' \text{ and } \theta' \sim \theta''. \tag{59.14}$$

($\Longleftarrow$) Suppose $\theta \sim \theta'$. Utilize (59.2) to invert the matrix to get

$$\theta' - \frac{-u\theta + s}{t\theta - r}, \tag{59.15}$$

because

$$\begin{pmatrix} r & s \\ t & u \end{pmatrix}^{-1} = \begin{pmatrix} -u & s \\ t & -r \end{pmatrix}. \tag{59.16}$$

Note that

$$q\theta - p = q\left(\frac{r\theta' + s}{t\theta' + u}\right) - p = \frac{q'\theta' - p'}{t\theta' + u}. \tag{59.17}$$

where

$$q' = qr - pt, \text{ and } p' = -qs + pu. \tag{59.18}$$

(Note: We may case this in matrix form as

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} r & -t \\ -s & u \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}. \tag{59.19}$$

Solving for p, q in terms of p', q':

$$\pm q = q'u + p't, \text{ and } \pm p = q's + p'r. \tag{59.20}$$

Next, rewrite the first equality in (59.18) as

$$q' = q(r - t\theta) + t(q\theta - p). \tag{59.21}$$

Without loss of generality, let

$$r - t\theta > 0. \tag{59.22}$$

If

$$|t(q\theta - p)| < r - t\theta \iff |q\theta - p| < \frac{r - t\theta}{|t|}, \tag{59.23}$$

then

$$\operatorname{sgn} q' = \operatorname{sgn} q. \tag{59.24}$$

The usefulness of (59.23) is that the bound is *uniform*. Let p_n/q_n and p_{n+1}/q_{n+1} be two successive best approximations (convergents) to θ and p'_n/q'_n and p'_{n+1}/q'_{n+1} be defined via (59.18). Then:

Claim: p'_n/q'_n and p'_{n+1}/q'_{n+1} are successive best approximations to θ' if n is large enough.

Proof: Observe that if n is large enough, then $(p, q) := (p_n, q_n)$ and $(r, s) = (p_{n+1}, q_{n+1})$ will satisfy (59.23). Therefore $\operatorname{sgn} q_n = \operatorname{sgn} q'_n > 0$ and $\operatorname{sgn} q_{n+1} = \operatorname{sgn} q'_{n+1} > 0$. In addition,

$$(p_{n+1} - p_n, q_{n+1} - q_n) =: (p, q)$$

also satisfies (59.23) for large n. Hence,

$$\operatorname{sgn} q_{n+1} - q_n = \operatorname{sgn} q'_{n+1} - q'_n < 0. \tag{59.25}$$

And so

$$0 < q'_n < q'_{n+1}. \ \square \tag{59.26}$$

Now, from (59.17)

$$|q'_n\theta' - p'_n| = |t\theta' + u||q_n\theta - p_n| > |t\theta' + u||q_{n+1}\theta - p_n| = |q'_{n+1}\theta' - p_{n+1'}|. \tag{59.27}$$

Suppose there are integers x', y' such that

$$0 < y' < q'_{n+1} \text{ and } y'\theta' - x'| \leq |q'_n\theta' - p'_n|. \tag{59.28}$$

We need to show that $(x', y') = (p'_n, q'_n)$.

Let (x, y) correspond to (x', y'). Thus,

$$|y\theta - x| \leq |q_n\theta - p_n|$$

by the principles underlying (59.27) and (59.19). Therefore

$$|y\theta - x| \leq \frac{r - t\theta}{2|t|}$$

for large enough n. Similarly

$$|q_{n+1}\theta - p_{n+1}| \leq \frac{r - t\theta}{2|t|}.$$

Thus, $(p, q) = (q_{n+1} - y, p_{n+1} - x)$ satisfies (59.23).

Therefore, (x, y) satisfies $0 < y < q_{n+1}$ and $|y\theta - x| \leq |q_n\theta - p_n|$, which forces

$$(x, y) = (p_n, q_n),$$

because p_n/q_n is a best approximation. Thus, by the transformation $(x', y') = (p'_{n+1}, q'_{n+1})$.

Thus, p'_n/q'_n and p'_{n+1}/q'_{n+1} are successive best approximations eventually. Thus p'_{n+1}/q'_{n+1} and p'_n/q'_n are successive convergents to θ'. Lastly, observe that

$$q'_{n+1} = rq_{n+1} - tp_{n+1} = r(c_nq_n + q_{n-1}) - t(c_np_n + p_{n-1}) = c_n(rq_n - tp_n) + (rq_{n-1} - tp_{n-1}) = c_nq'_n + q'_{n-1}. \tag{59.29}$$

This is the same recurrence that satisfies q_{n+1}. (Note: even though we're using the same index n, we don't claim that the convergent p_n/q_n to θ and its corresponding p'_n/q'_n to θ' happen at the same indices.) This proves the theorem. □

8.3 The Markov Constant of a Real Number

Given an irrational θ, define

$$\nu(\theta) := \liminf_{q \in \mathbb{Z}^+} q\|q\theta\| = \liminf_{n\to\infty} q_n\|q_n\theta\|, \tag{60.1}$$

where q_n is the denominator of the n^{th} best approximation to θ.

Dirichlet's Theorem implies that $0 \le \nu(\theta) \le 1$. Since one out of every pair of consecutive convergents to θ satisfies

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{2q^2},$$

we have the improved inequality $0 \le \nu(\theta) \le 1/2$.

Define the Markov Constant of θ as

$$M(\theta) = \frac{1}{\nu(\theta)}. \tag{60.2}$$

Thus, $M(\theta) = \infty$ if $\nu(\theta) = 0$. If $M(\theta) < \infty$, then

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{(M(\theta) - \epsilon)q^2} \tag{60.3a}$$

has infinitely many solutions, and

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{(M(\theta) + \epsilon)q^2} \tag{60.3b}$$

has only a finite number of solutions.

$M(\theta) = \infty$ implies that there exists $\lambda_n \to \infty$ such that

$$\left|\theta - \frac{p_n}{q_n}\right| < \frac{1}{\lambda_n q_n^2}. \tag{60.4}$$

Recall that if $\theta = \langle a_0, a_1, \cdots, a_n\rangle$, then $q_{n+1} = a_n q_n + q_{n-1}$, and so

$$\left|\theta - \frac{p}{q}\right| < \left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| = \frac{1}{q_n q_{n+1}} < \frac{1}{a_n q_n^2}. \tag{60.5}$$

Thus we have,
Theorem 60: If the partial quotients a_n are unbounded, then $M(\theta) = \infty$.

Corollary: $M(e) = \infty$.

Unsolved Problem: What is $M(\pi)$?

Theorem 61: Let θ and θ' be equivalent irrationals. Then, $\nu(\theta) = \nu(\theta')$.

Proof: Write

$$\theta = \frac{r\theta' + s}{t\theta' + u} \quad \text{from 59.2,}$$

with $r, s, t, u \in \mathbb{Z}$ and $|ru - ts| = 1$. Suppose $\kappa > 0$ is such that

$$q|q\theta - p| < \kappa \tag{61.1}$$

has infinitely many solutions. Define (p', q') using (p, q) as before. Then,

$$q' = qr - pt, \quad p' = -qs + pu \quad \text{from (59.18).}$$

From this, we get that

$$q\theta' - p' = pm\frac{q\theta - p}{r - t\theta}, \tag{61.2}$$

where the $\pm$ coincides with $\text{sgn}(ru - st)$. We have previously used the relation

$$q' = q(r - t\theta) - t(q\theta - p),$$

from (59.2). Combining (59.21) and (61.2), we get the following:

$$\begin{aligned} q'|q\theta' - p'| &\leq q|q\theta - p| + \frac{|t|(q\theta - p)^2}{|r - t\theta|}. \\ &< \kappa + \frac{|t||q\theta - p|^2}{|r - t\theta|} \\ &< \kappa + \frac{|t|}{|r - t\theta|}\frac{\kappa^2}{q^2}. \end{aligned} \tag{61.3}$$

By letting $q \to \infty$ (guaranteed because (61.1) has infinitely many solutions), we conclude that if $\kappa' > \kappa$, then

$$q'|q'\theta' - p'| < \kappa' \tag{61.4}$$

has infinitely many solutions. Since (61.4) has infinitely many solutions for every $\kappa' > \kappa$, we conclude that

$$\nu(\theta') \leq \nu(\theta).$$

Since $\theta \sim \theta'$ is a symmetric relation, we should also have that $\nu(\theta) \leq \nu(\theta')$, and thus

$$\nu(\theta) = \nu(\theta'). \ \square$$

Natural Question: Is the converse of **Theorem 61** true? Well it's definitely false for $\nu(\theta) = 0$, but even in the non-zero case, there uncountable many irrationals θ with $\nu(\theta) = 3$, even though there are only countable many irrationals equivalent to any given irrational θ.

Theorem 62: (Hurwitz) Let θ be irrational. Then, there exist infinitely many rationals p/q such that

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{\sqrt{5}q^2}. \tag{62.1}$$

The constant $\sqrt{5}$ is best possible in the sense that for every $\epsilon > 0$

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{(\sqrt{5} + \epsilon)q^2} \tag{62.2}$$

has only a finite number of solutions if

$$\theta \sim \frac{1 + \sqrt{5}}{2} = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{\ddots}}} =: \alpha.$$

Hence $M(\alpha) = \sqrt{5}$.

Proof: We focus on (p, q) being (p_n, q_n), the best approximations to θ. Put

$$A_n := q_n\|q_n\theta\|. \tag{62.3}$$

We know that

$$q_{n+1}\|q_n\theta\| + q_n\|q_{n+1}\theta\| = 1. \tag{62.4}$$

Define

$$\lambda := \frac{q_{n-1}}{q_n} \text{ and } \mu := \frac{q_{n+1}}{q_n}. \tag{62.5}$$

From (62.4), we deduce

$$\frac{q_n}{q_{n-1}}q_{n-1}\|q_{n-1}\theta\| + \frac{q_{n-1}}{q_n}q_n\|q_n\theta\| = 1, \tag{62.6}$$

which yields

$$\frac{A_{n-1}}{\lambda} + \lambda A_n = 1$$

and therefore

$$\lambda^2 A_n - \lambda + A_{n-1} = 0. \tag{62.7}$$

Similarly (62.4) yields

$$\mu^2 A_n - \mu + A_{n+1} = 0. \tag{62.8}$$

Note that

$$\mu - \lambda = \frac{q_{n+1} - q_n}{q_n} =: a_n. \tag{62.9}$$

From (62.7) and (62.8), we get

$$(\lambda^2 - \mu^2)A_n + (\mu - \lambda) + A_{n-1} - A_{n+1} = 0,$$

and therefore,

$$a_n(\lambda + \mu)A_n = a_n + A_{n-1} - A_{n+1}. \tag{62.10}$$

On the other hand

$$(\lambda^2 - \mu^2)A_n - (\lambda + \mu) + A_{n-1} + A_{n+1} = 0. \tag{62.11}$$

Multiply both sides of (62.11) by $2a_n^2A_n$ to get

$$qa_n^2(\lambda^2 - \mu^2)A_n^2 - 2a_n^2A_n(\lambda + \mu) + 2a_n^2A_n(A_{n-1} + A_{n+1}) = 0. \tag{62.12}$$

By (62.10)

$$-2a_n^2A_n(\lambda + \mu) = -2a_n(a_n + A_{n-1} + A_{n+1}. \tag{62.13}$$

Therefore

$$\begin{aligned} 2a_n^2(\lambda^2 + \mu^2)A_n^2 - (a_n + A_{n-1} + A_{n+1})^2 &= 2a_n^2(\lambda^2 + \mu^2)A_n^2 - a_n^2A_n^2(\lambda + \mu)^2 \\ &= a_n^2A_n^2(2\lambda^2 + 2\mu^2 - \lambda^2 - \mu^2 - 2\lambda\mu) \\ &= a_n^2A_n^2(\lambda - \mu)^2 \\ &= a_n^4A_n^2. \end{aligned} \tag{62.14}$$

Thus we may rewrite (62.12) as

$$a_n^4A_n^2 + (a_n + A_{n-1} + A_{n+1})^2 - 2a_n(a_n + A_{n-1} - A_{n+1}) + 2a_n^2A_n(A_{n-1} + A_{n+1}) = 0. \tag{62.15}$$

Observe that

$$\begin{aligned} (a_n + A_{n-1} - A_{n+1})^2 - 2a_n(a_n + A_{n-1} - A_{n+1}) &= a_n - (a_n + A_{n-1} - A_{n+1})^2 - a_n^2 \\ &= (A_{n-1} - A_{n+1})^2 - a_n^2. \end{aligned} \tag{62.16}$$

Combining (62.16) with (62.15) yields

$$a_n^4 A_n^2 + (A_{n-1} + A_{n+1})^2 - a_n^2 + 2a_n^2 A_n (A_{n-1} + A_{n+1}) = 0. \tag{62.17}$$

Canceling a_n^2 throughout, we rewrite (62.17) as

$$a_n^2 A_n^2 + 2A_n(A_{n-1} + A_{n+1}) = 1 - a_n^{-2}(A_{n-1} - A_{n+1})^2 \leq 1. \tag{62.18}$$

Now,

$$\text{LHS}(62.18) \geq (a_n^2 + 4) \cdot \min(A_{n-1}^2, A_n^2, A_{n+1}^2) \tag{62.19}$$

with the equality strict if $A_{n-1} = A_n = A_{n+1}$, and so we have

$$(a_n^2 + 4) \cdot \min(A_{n-1}^2, A_n^2, A_{n+1}^2) \leq 1. \tag{62.20}$$

Since for all n $a_n \geq 1$, in order for (62.20) to hold, we must have either

$$\min(A_{n-1}, A_n, A_{n+1}) < \frac{1}{\sqrt{5}}, \text{ or,} \tag{62.21}$$

$$\min(A_{n-1}, A_n, A_{n+1}) = \frac{1}{\sqrt{5}}, \text{ in which case } a_n = 1. \tag{62.22}$$

Note that (62.22) cannot hold. This forces $A_{n-1} = A_n = A_{n+1}$, and this forces λ and μ to be irrational from (62.7) and (62.8), which contradicts our assumption that λ and μ are rational. So, (62.21) holds, and thus one out of every three consecutive convergents satisfies the **Hurwitz Inequality**. So, there are infinitely many.

Now we show that this is best possible. Let

$$\alpha := \frac{1 + \sqrt{5}}{2}. \tag{62.23}$$

Then, α is a root of $x^2 - x - 1 = 0$. Let $p/q \neq \alpha$. Consider

$$f\left(\frac{p}{q}\right) - f(\alpha) = \left(\frac{p}{q} - \alpha\right) f'(\xi), \tag{62.24}$$

for some ξ between α and p/q, by the Mean Value Theorem.

Note that $f(\alpha) = 0$. Thus by (62.24), we have

$$\left| f\left(\frac{p}{q}\right) \right| = \left| a - \frac{p}{q} \right| |f'(\xi)|. \tag{62.25}$$

But then

$$f'(\xi) = 2\xi - 1.$$

So, let $p/q \to \alpha$, which implies $\xi \to \alpha$, and so

$$f'(\xi) \to 2\alpha - 1 = \sqrt{5}. \tag{62.26}$$

Thus by (62.25) and (62.26),

$$\left| \alpha - \frac{p}{q} \right| = \frac{\left| f\left(\frac{p}{q}\right) \right|}{|f'(\xi)|} = \frac{\left| p^2 - pq - q^2 \right|}{q^2 |f'(\xi)|} \geq \frac{1}{|f'(\xi)| q^2},$$

because the numerator is not equal to zero. Since $f'(\xi) \to \sqrt{5}$, we have that $\sqrt{5}$ as the best constant for $(1 + \sqrt{5})/2$. Hence $\sqrt{5}$ is $M(\theta)$ for $\theta \sim \alpha$. $\square$

Remark: The irrationals equivalent to α are those for which $a_n = 1$ for all $n \geq N$. Let us call this set of irrationals E_α. Thus if $\theta \in \mathscr{I} \setminus E_\alpha$, then infinitely many a_n are ≥ 2. If any $a_n \geq 2$, then $a^2 + 4 \geq 8$ for that n. The proof of **Theorem 62** shows that for any n such that $a_n \geq 2$, we must have $\min(A_{n-1}, A_n, A_{n+1}) < 1/\sqrt{8}$. So this yields:

Theorem 63: If $\theta \in \mathscr{I} \setminus E_\alpha$, then

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{\sqrt{8}q^2}$$

has infinitely many solutions.

Remark: Is this best possible? Consider

$$2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}} =: \beta.$$

Then, $\beta = 1 + \sqrt{2}$. This β is a root of $g(x) = x^2 - 2x - 1 = 0$ and $g'(x) = 2x - 2 = 2\sqrt{2} = \sqrt{8}$. So the above argument will show

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{(\sqrt{8} + \epsilon)q^2}$$

holds only finitely often. Thus $M(\theta) = \sqrt{8}$ for all irrationals equivalent to $1 + \sqrt{2}$.

Consider

$$\theta \in \mathscr{I} \setminus E_\alpha \setminus E_\beta.$$

We have a sequence $E_\alpha, E_\beta, E_\gamma$, etc, with $M(\alpha) < M(\beta) < M(\gamma) < \cdots$.

Theorem: The set of irrationals θ for which $M(\theta) = 3$ is uncountable.

Remark: We know two estimates for $|\theta - P_n/Q_n|$:

$$\left|\theta - \frac{P_n}{Q_n}\right| < \left|\frac{P_n}{Q_n} - \frac{P_{n+1}}{Q_{n+1}}\right| = \frac{1}{Q_n Q_{n+1}}. \tag{65.1}$$

$$\theta - \frac{P_n}{Q_n} = \frac{(-1)^n}{Q_n(\theta_{n+1}Q_n + Q_{n-1})}. \tag{65.2}$$

Thus, in order to understand $Q_n\|Q_n\theta\|$, we need to understand Q_{n+1}/Q_n.

Theorem 65: Let

$$\theta = a_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\cdots,$$

and let P_n/Q_n be the n^{th} convergent. Then

$$\frac{P_n}{P_{n-1}} = b_n + \frac{a_n}{b_{n-1}+}\frac{a_{n-1}}{b_{n-1}+}\cdots\frac{a_1}{b_1+}\frac{a_1}{a_0},$$

$$\frac{Q_n}{Q_{n-1}} = b_n + \frac{a_n}{b_{n-1}+}\frac{a_{n-1}}{b_{n-2}+}\cdots\frac{a_2}{b_2+}\frac{a_2}{b_1}.$$

Proof: The recurrence

$$P_n = b_n P_{n-1} + a_n P_{n-2}$$

implies

$$\frac{P_n}{P_{n-1}} = b_n + \frac{a_n P_{n-1}}{P_{n-1}} = b_n + \frac{a_n}{P_{n-1}/P_{n-2}}.$$

Iterate this. □

In the case of simple continued fractions,

$$\frac{P_n}{P_{n-1}} = \langle a_n, a_{n-1}, \cdots, a_1, a_0\rangle, \text{ for } n \geq 1,$$

$$\frac{Q_n}{Q_{n-1}} = \langle a_n, a_{n-1}, \cdots, a_2, a_1\rangle, \text{ for } n \geq 2.$$

Therefore,

$$Q_n\|Q_n\theta\| = \frac{1}{Q_{n+1} + (Q_{n-1}/Q_n)} = \frac{1}{\langle a_{n+1}, a_{n+2}, \cdots\rangle + \langle 0, a_n, a_{n-1}, \cdots, a_1\rangle}. \tag{65.3}$$

Theorem 66: If $\theta = \langle a_0, a_1, \cdots\rangle$, then

$$\nu(\theta) = \liminf\left\{\frac{1}{\langle a_{n+1}, a_{n+2}, \cdots\rangle + \langle 0, a_n, a_{n-1}, \cdots, a_1\rangle}\right\}, \tag{66.1}$$

and so,

$$M(\theta) = \limsup\{\langle a_{n+1}, a_{n+2}, \cdots\rangle + \langle 0, a_n, a_{n-1}, \cdots, a_1\rangle\}. \tag{66.2}$$

Chapter 9

Farey Fractions

Definition: The Farey Sequence of order n, denoted by $\mathbb{F}_n$ is the collection of all reduced fractions p/q in ascending order, with $q \leq n$. Clearly,

$$\mathbb{F}_n \cap [m, m+1) = (\mathbb{F}_n \cap [0,1)) + m. \tag{67.1}$$

Hence, we focus only on the half-open unit interval $[0, 1)$.

Example:

$$\mathbb{F}_5 \cap [0,1] = \frac{0}{1}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1}.$$

Remark:

$$|\mathbb{F}_n \cap [0,1]| = 1 + \varphi(1) + \varphi(2) + \cdots + \varphi(n), \tag{67.2}$$

where φ is the Euler function.

Theorem 67: Let p/q and r/s be two rationals satisfying

$$|ps - rq| = 1. \tag{67.3}$$

Then, p/q and r/s are consecutive fractions in $\mathbb{F}_n$ for n satisfying

$$\max(q, s) \leq n < q + s. \tag{67.4}$$

Proof: Without loss of generality, assume that

$$\frac{p}{q} < \frac{r}{s} \text{ and hence } qr - ps = 1. \tag{67.5}$$

Consider the function

$$f(t) = \frac{p + rt}{q + st}, \text{ for } t \geq 0. \tag{67.6}$$

Note that f is strictly increasing because

$$\begin{aligned} f'(t) &= \frac{(q+st)r - (p+rt)s}{(q+st)^2} \\ &= \frac{qr - ps}{(q+st)^2} \\ &= \frac{1}{(q+st)^2} > 0, \end{aligned}$$

and therefore

$$f : [0, \infty) \to \left[\frac{p}{q}, \frac{r}{s}\right) \tag{67.7}$$

is one-to-one and onto.

Note also that

$$t \in \mathbb{Q} \iff f(t) \in \mathbb{Q}$$

because

$$\begin{pmatrix} p & r \\ q & s \end{pmatrix}^{-1}$$

is a unimodular matrix as well.

Thus, f maps non-negative rationals in a one-to-one and onto fashion to the rationals in $\left[\frac{p}{q}, \frac{r}{s}\right)$.

Let $u/v \in \mathbb{Q}$, with $\gcd(u, v) = 1$, and consider

$$f\left(\frac{u}{v}\right) = \frac{p + r\frac{u}{v}}{q + s\frac{u}{v}} = \frac{pv + ru}{qv + su} =: \frac{a}{b}, \tag{67.8}$$

with $a := pv + ru$ and $b := qv + su$.

Observe that

$$q(pv + ru) - p(qv + su) = u(qr - sp) = u. \tag{67.9}$$

Similarly,

$$s(pv + ru) - r(qv + su) = v(sp - rq) = -v. \tag{67.10}$$

Since $\gcd(u, v) = 1$, there exists integers A and B such that

$$Au + Bv = 1. \tag{67.11}$$

Therefore, from (67.9), (67.10), and (67.11), we see that there exist integers $A', B' \in \mathbb{Z}$ such that $A'a + B'b = 1$ if and only if $\gcd(a, b) = 1$. Thus f takes reduced rationals to reduced rationals.

Clearly, $\max(q, s) \leq n$, which is equivalent to $\frac{p}{q}, \frac{r}{s} \in \mathbb{F}_n$.

According to the above discussion, all reduced fractions in $\left(\frac{p}{q}, \frac{r}{s}\right)$ are given by $\frac{pv + ru}{qv + su}$ with $\gcd(u, v) = 1$ and $u, v > 0$.

This means that the denominators of any fraction in the range $\left(\frac{p}{q}, \frac{r}{s}\right)$ are at least as big as $q + s$. Thus the theorem is proved. □

Algorithm to generate $\mathbb{F}_{n+1}$ from $\mathbb{F}_n$: Consider all fractions of the form $\frac{p+r}{q+s}$ from consecutive $\frac{p}{q}, \frac{r}{s} \in \mathbb{F}_n$, and choose only those with $q + s = n + 1$. The result combined with $\mathbb{F}_n$ is $\mathbb{F}_{n+1}$.

Definition: The fraction $\frac{p+r}{q+s}$ is called the mediant of $\frac{p}{q}$ and $\frac{r}{s}$.

Lemma: (a property of mediants) The mediant of two fractions lies between them, i.e.,

$$\frac{p}{q} < \frac{r}{s} \Longrightarrow \frac{p}{q} < \frac{p+r}{q+s} < \frac{r}{s}.$$

Moreover, the "cross-difference" is preserved by the process of taking mediants, i.e.,

$$qr - ps = q(p + r) - p(q + s) = (q + s)r - (p + r)s.$$

Therefore, if $qr - ps = 1$, then so are the two cross-differences, and then so the mediant is *reduced* as well.

Theorem 68: Let $\frac{p}{q}$ and $\frac{r}{s}$ be two consecutive fractions in some $\mathbb{F}_n$. Then $|ps - qr| = 1$.

Proof: In $\mathbb{F}_1$, all elements are integers, and this is clear. To construct $\mathbb{F}_{n+1}$ from $\mathbb{F}_n$, we use the mediants as in the above **Algorithm to generate** $\mathbb{F}_{n+1}$ **from** $\mathbb{F}_n$. But the mediant preserves cross-differences by the above **Lemma**. Hence the theorem is true. □

Exercise: If $\frac{p}{q}$, $\frac{p'}{q'}$, $\frac{p''}{q''}$ are three consecutive fractions in some $\mathbb{F}_n$, then $\frac{p + p''}{q + q''} = \frac{p'}{q'}$. Warning: the thing that makes this non-trivial is that $\frac{p + p''}{q + q''}$ need not be reduced!

Theorem 69: Let θ be irrational and $\theta \in \left(\frac{p}{q}, \frac{r}{s}\right)$, with $qr - ps = 1$. Then, one of the three fractions $\frac{p}{q}$, $\frac{p + r}{q + s}$, $\frac{r}{s}$ will satisfy

$$\left|\theta - \frac{a}{b}\right| < \frac{1}{\sqrt{5}b^2}.$$

Proof: (version 1) For a rational a/b, define the interval

$$I_\lambda\left(\frac{a}{b}\right) := \left(\frac{a}{b} - \frac{1}{\lambda b^2}, \frac{a}{b} + \frac{1}{\lambda b^2}\right). \tag{69.1}$$

If λ is such that

$$I_\lambda\left(\frac{p}{q}\right) \cup I_\lambda\left(\frac{r}{s}\right) \supseteq \left[\frac{p}{q}, \frac{r}{s}\right]. \tag{69.2}$$

Then,

$$\left|\theta - \frac{a}{b}\right| < \frac{1}{\lambda b^2} \tag{69.3}$$

for either $\frac{p}{q} = \frac{a}{b}$ or $\frac{r}{s} = \frac{a}{b}$.

Note that (69.2) holds if and only if

$$\frac{1}{\lambda q^2} + \frac{1}{\lambda s^2} > \frac{r}{s} - \frac{p}{q} = \frac{1}{qs}, \tag{69.4}$$

which is equivalent to

$$\lambda \leq \frac{qs}{q^2} + \frac{qs}{s^2} = \frac{s}{q} + \frac{q}{s}. \tag{69.5}$$

But,

$$g(x) = x + \frac{1}{x}, \tag{69.6}$$

and hence

$$\lambda \leq g\left(\frac{s}{q}\right) = g\left(\frac{q}{s}\right). \tag{69.7}$$

Without loss of generality, let

$$\frac{p}{q} < \theta < \frac{p + r}{q + s}. \tag{69.8}$$

We now want

$$I_\lambda\left(\frac{p}{q}\right) \cup I_\lambda\left(\frac{p+r}{q+s}\right) \supseteq \left[\frac{p}{q}, \frac{p+r}{q+s}\right]. \tag{69.9}$$

Thus we want

$$\lambda \leq g\left(\frac{q+s}{q}\right) = g\left(1+\frac{s}{q}\right),$$

and therefore,

$$\lambda \leq \min_x\{\max(g(x), g(1+x))\}.$$

Then, for such a λ, we would have the theorem hold. □

Proof: (version 2)

Lemma: There are no positive integers q and s such that

$$\frac{1}{qs} \geq \frac{1}{\sqrt{5}}\left(\frac{1}{q^2}+\frac{1}{s^2}\right) \tag{69.21}$$

and

$$\frac{1}{q(q+s)} \geq \frac{1}{\sqrt{5}}\left(\frac{1}{q^2}+\frac{1}{(q+s)^2}\right) \tag{69.22}$$

Proof: Suppose (69.21) and (69.22) both hold. Then

$$q^2+s^2-\sqrt{5}qs \leq 0, \tag{69.23a}$$

and

$$(q+s)^2+q^2-\sqrt{5}(q+s)q \leq 0. \tag{69.23b}$$

Thus,

$$2q^2+s^2+2sq-\sqrt{5}q^2-\sqrt{5}sq \leq 0.$$

and therefore

$$(2-\sqrt{5})(q^2+sq)+s^2 \leq 0. \tag{69.24}$$

Adding the inequalities:

$$(3-\sqrt{5})q^2+2(1-\sqrt{5})sq+2s^2 \leq 0.$$

Multiplying by two:

$$6-2\sqrt{5})q^2+4(1-\sqrt{5})sq+4s^2 \leq 0. \tag{69.25}$$

Observe that

$$(1-\sqrt{5})^2 = 6-2\sqrt{5} \tag{69.26}$$

and so we have

$$(1-\sqrt{5})q+2s)^2 \leq 0,$$

but the left hand side is not zero, so this is a contradiction. □

Let θ be irrational, and let

$$\frac{p}{q} < \theta < \frac{r}{s},$$

with $rq-ps=1$. Without loss of generality, we have that

$$\frac{p}{q} < \theta < \frac{p+r}{q+s} < \frac{r}{s}. \tag{69.27}$$

Assume that the Theorem is false. Then from (69.27):

$$\begin{aligned}\theta - \frac{p}{q} &\geq \frac{1}{\sqrt{5}q^2},\\ \frac{r}{s} - \theta &\geq \frac{1}{\sqrt{5}s^2},\\ \frac{p+r}{q+s} - \theta &\geq \frac{1}{\sqrt{5}(1+s)^2}.\end{aligned} \tag{69.28}$$

Thus,

$$\frac{r}{s} - \frac{p}{q} = \frac{1}{sq} \geq \frac{1}{\sqrt{5}}\left(\frac{1}{q^2} + \frac{1}{s^2}\right)$$

and

$$\frac{p+r}{q+s} - \frac{p}{q} = \frac{1}{q(q+s)} \geq \frac{1}{\sqrt{5}}\left(\frac{1}{q^2} + \frac{1}{(q+s)^2}\right).$$

This contradicts the lemma above, so the theorem holds. □

Chapter 10

Transcendental Numbers

10.1 Khintchin's Metric Theorems

Let $f(q)$ be increasing over $q \in \mathbb{Z}^+$, with $f(1) \geq 2$ and let $f(q) \to \infty$ as $q \to \infty$. Then, either

$$\sum \frac{1}{f(q)} < \infty, \tag{70.1}$$

or

$$\sum \frac{1}{f(q)} = \infty. \tag{70.2}$$

Theorem 70: (Khintchin's First Theorem) Let f be as above, and let (70.1) hold. Let S be the set of all reals such that $\left|\theta - \frac{p}{q}\right| < \frac{1}{qf(q)}$ has infinitely many solutions. Then, S has measure zero.

Proof: Let $S_{0,1} := S \cap [0,1]$. Note that $S_{m,m+1} = S \cap [m, m+1] = S_{0,1} + m$. So, it suffices just to show that the measure of $S_{0,1}$ is zero (because a countable union of sets of measure zero must have measure zero).

In view of the convergence, if $\epsilon > 0$ is given, then there exists n such that

$$\sum_{q>N} \frac{1}{qf(q)} < \epsilon. \tag{70.3}$$

Define

$$I_q(f) := \left(0,\ \frac{1}{qf(q)}\right) \cup \left(\frac{1}{q} - \frac{1}{qf(q)},\ \frac{1}{q} + \frac{1}{qf(q)}\right) \cup \left(\frac{2}{1} - \frac{1}{qf(q)},\ \frac{2}{q} + \frac{1}{qf(q)}\right) \cup \cdots \cup \left(1 - \frac{1}{qf(q)},\ 1\right).$$

If $\theta \in S_{0,1}$, then $\theta \in I_q(f)$ for infinitely many q. Therefore, $S_{0,1} \subseteq \bigcup_{q>N} I_q(f)$, for each N. Hence

$$\mu(S_{0,1}) \leq \sum_{q>N} \mu(I_q(f)) = \sum_{q>N} \frac{2}{f(q)} < 2\epsilon,$$

with $\epsilon > 0$. Thus, $\mu(S_{0,1}) = 0$ as claimed. □

Theorem 71: (Khintchin's Next Theorem) Let $f(q)$ be as above and let (70.2) hold. Then, almost all reals θ satisfy $\left|\theta - \frac{p}{q}\right| < \frac{1}{qf(q)}$ for infinitely many p/q.

Proof: Very difficult.

Remarks: The set of irrationals satisfying this is of measure 0:

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^2 \log(q)^2}$$

Almost all irrationals satisfy this:

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^2 \log(q)}$$

10.2 The First Transcendental Number

Theorem 72: (Liouville's Theorem) Let α be an algebraic number of degree n. Then, there exists $c(\alpha)$ such that

$$\left|\alpha - \frac{p}{q}\right| > \frac{c(\alpha)}{q^2},$$

if $\deg \alpha \neq 0$.

Proof: Clearly, the theorem holds for the rational case ($\alpha = 1$). Now let α be irrational, i.e., $\deg \alpha \geq 2$. Let $P(x)$ be the minimal polynomial of α, with $P(x) \in \mathbb{Z}[x]$, $P(\alpha) = 0$, and P is irreducible. Now, by the Mean Value Theorem, we have that

$$P(\alpha) - P\left(\frac{p}{q}\right) = P'(\xi)\left(\alpha - \frac{p}{q}\right) \tag{72.1}$$

for some ξ lying between α and p/q. We're assuming that $p/q \in (\alpha - 1, \alpha + 1)$. Let

$$\kappa := \max_{\xi \in [\alpha-1, \alpha+1]} |P'(\xi)|. \tag{72.2}$$

Therefore, from (72.1) and (72.2), we get

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{|P(\alpha - P(p/q)|}{|P'(\xi)|} \geq \frac{|P(p/q)|}{\kappa} \tag{72.3}$$

because $P(\alpha) = 0$.

Next, since P is irreducible over $\mathbb{Q}$, we have that $P(p/q) \neq 0$. Hence

$$q^n P(p/q) \in \mathbb{Z} \setminus \{0\}. \tag{72.4}$$

Therefore,

$$|q^n P(p/q)| \geq 1. \tag{72.5}$$

Thus, from (72.3) and (72.5)

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{1}{\kappa q_n} \text{ if } \left|\alpha - \frac{p}{q}\right| < 1.$$

If $|\alpha - p/q| \geq 1$, then the inequality holds with $c(\alpha) = 1$, so choose $c(\alpha) := \min\left(1, \frac{1}{\kappa}\right)$. □

Theorem 73: The number λ defined by

$$\lambda := \sum_{n=0}^{\infty} \frac{1}{10^{n!}}$$

is not algebraic.

Proof: Notice that

$$\left|\lambda - \sum_{m=0}^{n} \frac{1}{10^{m!}}\right| \leq \frac{2}{10^{(n+1)!}},$$

i.e.,

$$\left|\lambda - \frac{p_n}{q_n}\right| \leq \frac{4}{q^{n+1}}.$$

Chapter 11

Irrationality Type and Measure

11.1 Definitions

Definition: Let θ be irrational. We say that θ is of irrationality type $\leq \tau$ if for $\epsilon > 0$,

$$\left|\theta - \frac{p}{q}\right| > \frac{c(\epsilon)}{q^{\tau+\epsilon}}$$

for some $c(\epsilon)$ and for all rationals p/q. Thus,

$$\left|\theta - \frac{p}{q}\right| > \frac{1}{q^{\tau+\epsilon}}$$

has only a finite number of solutions, for every $\epsilon > 0$.

Definition: If $\tau_0 = \inf \tau$ such that θ is type $\leq \tau$, we say θ is of τ_0.

Definition: Let θ be irrational. We say that μ is an irrationality measure for θ if there exists a constant c such that

$$\left|\theta - \frac{p}{q}\right| > \frac{c}{q^{\mu}}$$

for all rationals p/q. This is an effective measurement. We need an effectively computable c.

Remark: Clearly, $\mu \geq \tau_0 = \tau(\theta)$.

Remark: By **Dirichlet's Theorem**,

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^2}$$

has infinitely many solutions. By **Liouville's Theorem**, if θ is a quadratic irrational, then

$$\left|\theta - \frac{p}{q}\right| > \frac{c}{q^2}$$

for $c = c(\theta)$. Thus, $\tau(\theta) = 2$ for all quadratic irrationals.

Theorem 70$*$: Almost all reals have irrationality type 2.

Proof: For each $k \in \mathbb{Z}^+$, consider the set S_k such that $\tau(\theta) < 2 + \dfrac{1}{k}$, where $S_k = \{\theta \mid \tau(\theta) > 2 + \frac{1}{k}\}$. Thus,

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^{1+(1/2)k}}$$

has infinitely many solutions. By **Khintchin's Theorem**, since

$$\sum_{q=1}^{\infty} \frac{1}{q^{1+(1/k)}} < \infty,$$

for each fixed $k \in \mathbb{Z}^+$, we have $\mu(S_k) = 0$. Thus,

$$\mathscr{S} := \bigcup_{k=1}^{\infty} S_k$$

has $\mu(\mathscr{S}) = 0$. But for $\theta \in \mathscr{I} \setminus \mathscr{S}$, we have $\tau(\theta) = 2$. But $\mathscr{I} - \mathscr{S}$ is almost all of $\mathbb{R}$. Hence the theorem holds. $\square$

11.2 The Thue-Siegel-Roth-Dyson Theorem

Remark: Liouville's Theorem implies that if α is algebraic, then $\tau(\alpha) \leq \deg \alpha$. Axel Thue realized that proving $\tau(\alpha) < \deg \alpha$ when $\deg \alpha \geq 3$ has important consequences.

Theorem T: (Thue, 1909, Crelle) Let α be algebraic. Then

$$\tau(\alpha) \leq \frac{n}{2} + 1$$

Let $n \geq 3$. Then, the Diophantine equation

$$x^n - Dy^n = k \tag{T_0}$$

has only a finite number of solutions. In contrast,

$$x^2 - Dy^2 = k$$

can have ∞'ly many solutions. This is Pell's Equation.

Proof: Rewrite (T_0) as

$$(x - \sqrt[n]{D}y)(x^{n-1} + x^{n-2}\sqrt[n]{D}y + \cdots + \sqrt[n]{D}^{n-1}y^{n-1}) = k. \tag{T_1}$$

We may view (T_1) as

$$\left| \sqrt[n]{D} - \frac{x}{y} \right| = O\left(\frac{1}{y^n}\right).$$

This would violate **Theorem T** if $y \to \infty$. $\square$

Remark: More generally, if $F(x, y)$ is a form of $\deg n \geq 3$, then $F(x, y) = m$ has only a finite number of solutions integers x, y.

In 1921, Siegel published the following theorem:
Theorem: (Siegel, 1921, Math Z.) $\tau(\alpha) \leq 2\sqrt{n}$. More precisely, $\tau(\alpha) \leq \min\limits_{1 \leq \beta \leq n-1} \left(\frac{n}{\beta + 1} + \beta \right)$.

This was followed by:
Theorem: (Dyson, 1947, Acta Mathematica) $\tau(\alpha) \leq \sqrt{2n}$.

These improvements seemed to suggest:
Theorem: (Roth, 1955, Mathematika) If α is algebraic, then $\tau(\alpha) = 2$.

11.3 Methods for Obtaining Irrationality Type and Measure

Theorem 74: Let K denote the set of rationals or an imaginary quadratic field. Let R be the ring of integers in K. Let $\theta \in \mathbb{C}$.

(a) (Asymptotic Version) Suppose there exists $Q > 1$, $E > 1$, and $p_n, q_n \in R$ satisfying

$$\left.\begin{array}{c} |q_n| \leq Q^{n(1+o(1))} \\ |q_n\theta - p_n| \leq E^{-n(1+o(1))} \end{array}\right\} \tag{74.1}$$

and

$$p_n q_{n+1} \neq q_n p_{n+1}. \tag{74.2}$$

Then, $\theta \notin K$. Moreover, given $\epsilon > 0$, there exists $b_0 = b_0(\epsilon)$ such that if $a, b \in R$, with $|b| > b_0(\epsilon)$, then

$$\left|\theta - \frac{a}{b}\right| > \frac{1}{b^{\tau+\epsilon}} \tag{74.3}$$

where

$$\tau = \frac{\log QE}{\log E} = 1 + \frac{\log Q}{\log E} \tag{74.4}$$

Thus θ is of type $\leq \tau$.

(b) (Effective Version) Suppose there exists $Q, E > 1$, $K_0 > 0$, $L_0 \geq 1/2$, and $p_n, q_n \in R$ satisfying (74.2) together with the following

$$|q_n| < k_0 Q^n \text{ and } |q_n\theta - p_n| \leq \ell_0 E^{-n} \tag{74.5}$$

Then, $\sigma \notin K$. Moreover, for all $a, b \in K$, we have that

$$\left|\theta - \frac{a}{b}\right| > \frac{c}{|b|^u} \tag{74.6}$$

where

$$\mu = \frac{\log QE}{\log E} = 1 + \frac{\log Q}{\log E} \tag{74.7a}$$

and

$$c = \frac{1}{2k_0 Q^{\left(2+\frac{\log(2\ell_0)}{\log E}\right)}} \tag{74.7b}$$

Thus μ is an irrationality measure for θ.

Proof of effective version: From (74.5), we see that

$$|q_n\theta - p_n| \to 0, \text{ as } n \to \infty.$$

Therefore,

$$|q_n\theta - p_n|^{-1} \to \infty, \text{ as } n \to \infty.$$

We have by **Dirichlet's Criterion** that $\theta \notin K$, because either $p_n/q_n \neq 0$ or $p_{n+1}/q_{n+1} \neq 0$.

Let m be the least positive integer such that

$$|q'_m\theta - p'_m|^{-1} \geq 2|b|, \text{ for all } m' \geq m. \tag{74.8}$$

Suppose that n is such that

$$\ell_0 E^{-n} \leq \frac{1}{2|b|} \iff 2|b|\ell_0 \leq E^n \iff n \geq \frac{\log(2|b|\ell_0)}{\log E} \tag{74.9}$$

Thus (74.5) would imply

$$|q_n\theta - p_n| \leq \frac{1}{2|b|}$$

for n satisfying (74.9).

Therefore,

$$|q_n\theta - p_n|^{-1} \geq 2|b| \tag{74.10}$$

Thus,

$$mleq\frac{\log(2|b|\ell_0)}{\log E} + 1 \tag{74.11}$$

Given $a, b \in R$, choose $n = m$ or $n = m + 1$ such that

$$aq_n \neq bp_n \tag{74.12}$$

(guaranteed by (74.2)).

Thus, (74.10) implies that

$$\begin{aligned} |q_n|\left|\theta - \frac{a}{b}\right| &\geq |q_n|\left(\left|\frac{p_n}{q_n} - \frac{a}{b}\right| - \left|\theta - \frac{p_n}{q_n}\right|\right) \\ &\geq \frac{1}{|b|} - \frac{1}{2|b|} \\ &\geq \frac{1}{2|b|} \quad (74.13) \end{aligned}$$

because of (74.9) and the fact that $0 \neq aq_n - bp_n \in R$. Thus

$$\begin{aligned} \left|\theta - \frac{a}{b}\right| &> \frac{1}{2|b||q_n|} \\ &\geq \frac{1}{2|b|k_0Q^{m+1}} \\ &\geq \frac{1}{2|b|k_0Q^{\frac{\log(2|b_0|\ell_0)}{\log E} + 2}} \\ &= \frac{Q^{-\left(2+\frac{\log(2\ell_0)}{\log E}\right)}}{2k_0} \cdot \frac{1}{|b|^{1+\frac{\log Q}{\log E}}}, \end{aligned}$$

which is the claim. □

Remark: It is usually cumbersome to verify (74.2). So, we will establish a result which gets rid of (74.2).

Theorem 75: (Alladi, 1980) Let $\theta \in \mathbb{R}$ and $p_n, q_n \in \mathbb{Z}$, such that

(i) $q_n \to \infty$ as $n \to \infty$,

(ii) $q_{n+1} = q_n^{1+o(1)}$.

(iii) For some $\lambda \in (0,1)$

$$\left|\theta - \frac{p_n}{q_n}\right| = \frac{1}{q_n^{(1+\lambda)(1+o(1))}}.$$

Then,

$$\left|\theta - \frac{p}{q}\right| > \frac{1}{q_n^{\left(1+\frac{1}{\lambda}\right)(1+o(1))}},$$

i.e., θ is irrational of type $\leq 1 + \dfrac{1}{\lambda}$.

Remark: Given (75.i), (75.ii), and (75.iii), we can extract a subsequence P_n/Q_n satisfying the same condition together with (74.2).

Observe that by (74.1)

$$\begin{aligned}\left|\theta - \frac{p_n}{q_n}\right| &< \frac{E^{-n(1+o(1))}}{q_n} \\ &\leq \frac{1}{q^{1+\frac{n\log E}{\log q_n}(1+o(1))}} \\ &\leq \frac{1}{q^{1+\frac{\log E}{\log Q}(1+o(1))}}.\end{aligned}$$

Thus, $\lambda = \dfrac{\log E}{\log Q}$, and so $\dfrac{\log Q}{\log E} = \dfrac{1}{\lambda}$ in Theorem 74.

Theorem 76: Let $0 \neq s \in \mathbb{Q}$. Then e^s is irrational of type 2.

Proof: Given s, define

$$a_r = \int_0^1 e^{sx} x^r dx, \tag{76.1}$$

for $r = 0, 1, 2, \ldots$. The a_r are given recursively by

$$a_0 = \int_0^1 e^{sx} dx = \frac{e^s - 1}{s} \tag{76.2}$$

and

$$a_r = \frac{e^{sx}}{s} x^r \Big|_0^1 - \frac{r}{s}\int_0^1 e^{sx} x^{r-1} dx = \frac{e^s}{s} - \frac{r}{s} a_{r-1}, \tag{76.3}$$

for $r \geq 1$. Thus by iteration:

$$a_r = \frac{(-1)^r \cdot r!}{s^{r+1}} \left\{ \left(1 - s + \frac{s^2}{2!} - \cdots + \frac{(-1)^r s^r}{r!}\right) e^s - 1 \right\} = (-1)^r (u_r e^s - v_r) \tag{76.4}$$

If $s = p/q$, then from (76.4), we get

$$p^{r+1} u_r \in \mathbb{Z} \text{ and } p^{r+1} v_r \in \mathbb{Z}. \tag{76.5}$$

Now, we bring in the Legendre polynomials:

$$P_n(x) = \frac{1}{n!}\frac{d^n\{x^n(1-x)^n\}}{dx^n} \tag{76.6}$$

Note

$$P_n(x) \in \mathbb{Z}[x], \quad \deg P_n = n. \tag{76.7}$$

Define

$$I_n = p^{n+1}\int_0^1 e^{sx}P_n(x)dx \tag{76.8}$$

This is an integral linear combination of the a_r's. Therefore

$$I_n = q_n e^s - p_n, \text{ for } q_n, p_n \in \mathbb{Z}.$$

An n-fold integration by parts (integrating $P_n(x)$ and differentiating e^{sx}) yields

$$I_n = \frac{p^{n+1}s^n}{n!}\int_0^1 e^{sx}x^n(1-x)^n dx \tag{76.9}$$

This says that $I_n \neq 0$ because the integrand is positive. Additionally,

$$\max_{0\leq x\leq 1} e^{sx}x^n(1-x)^n = o\left(\frac{1}{4^n}\right). \tag{76.10}$$

Hence

$$0 \neq I_n = q_n e^s - p_n = o\left(\frac{p_n}{q_n n! 4^n}\right) \xrightarrow{n\to\infty} 0 \tag{76.11}$$

So, $e^s \notin \mathbb{Q}$.

Note that e^{sx} varies from 1 to e^{-s} or e^s for $0 \leq x$‘1. We know that for any continuuous function f on a closed and bounded interval,

$$\|f\|_n \xrightarrow{n\to\infty} \|f\|_\infty.$$

Thus, we infer that

$$I_n = q_n e^s - p_n = \frac{p_n}{q_n n!}\left(\frac{1}{4}\right) = \frac{1}{(n!)^{1+o(1)}} \tag{76.12}$$

Next observe that

$$a_r \sim \frac{(-1)^r r!}{s^{r+1}}\left\{\left(e^{-s} - \frac{s^{r+1}}{(r+1)!}(-1)^{r+1}\right)e^s - 1\right\} = \frac{e^s}{r+1} \text{ as } r\to\infty. \tag{76.13}$$

Therefore,

$$u_r \sim \frac{r!e^{-s}}{s^{r+1}}. \tag{76.14}$$

Note that the coefficients of $P_n(x)$ alternate in sign. This, combined with the alternating signs in (76.4) imply that there is no cancellation of the q_n, i.e.

$$q_n = (n!)^{1+o(1)}. \tag{76.15}$$

Moreover,

$$q_{n+1} = q_n^{1+o(1)}. \tag{76.16}$$

Thus we have constructed $p_n, q_n \in \mathbb{Z}$ that satisfy

$$0 \neq \left| e^s - \frac{p_n}{q_n} \right| = \frac{1}{q_n^{2(1+o(1))}}. \tag{76.17}$$

So, apply **Theorem 75** with $\lambda = 1$. Thus, e^s is of irrationality $\leq 1 + \frac{1}{\lambda} = 2$, and thus since it is irrational, we have e^s of irrationality type $= 2$. $\square$

Remark: The irrationality of e^s for all $0 \neq s \in \mathbb{Q}$ implies the irrationality of $\log s$ for all $0 \neq s \in \mathbb{Q}$. However, the irrationality measure for e^s does not transfer to an irrationality measure for $\log s$. So, we need to approach $\log s$ directly. We need some tools to do this.

Lemma: Let $\ell_n := \text{lcm}\{1, 2, \ldots, n\}$. Then,

$$\ell_n = e^{n(1+o(1))}.$$

Proof: Recall the **Chebychev Function**:

$$\psi(n) = \sum_{n \leq x} \Lambda(n) = \sum_{p^\alpha \leq x} \log p. \tag{77.1}$$

Note that

$$\ell_n = \prod_p p^{\alpha_p(n)}, \tag{77.2}$$

where

$$\alpha_p(n) = \max\{\alpha \mid p^\alpha \leq n\}. \tag{77.3}$$

Thus,

$$\ell_n = \prod_{p^\alpha \leq n} p = e^{\psi(n)}. \tag{77.4}$$

The **Prime Number Theorem** is equivalent to

$$\psi(n) = n(1 + o(1)). \tag{77.5}$$

Hence the lemma. $\square$

Remark: For irrationality measures (effective), we need explicit upper bounds for $\psi(x)$ and other related functions. Such bounds were established by Rossen & Schoenfeld in the Illinois J. Math, 64-93 in 1962:

$$\ell_n < (2.826)^n, \text{ for all } n. \tag{77.6}$$

We now want to evaluate the size of $P_n(z)$ asymptotically. For this, define

$$\alpha(z) := \max\left\{ \left| \frac{\left(1 \pm \sqrt{1-z}\right)^2}{x} \right| \right\} \tag{77.7}$$

$$\beta(z) := \min\left\{ \left| \frac{\left(1 \pm \sqrt{1-z}\right)^2}{x} \right| \right\} \tag{77.8}$$

for $x \notin [1, \infty)$.

Lemma: For each $z \in \mathbb{C}$, the $P_n(x)$ satisfy

$$nP_n(z) + (2n-1)(2z-1)P_{n-1}(z) + (n-1)P_{n-2}(z) = 0.$$

Theorem 77: (Poincaré) Let $\{u_n\}$ be a sequence of complex numbers satisfying

$$a_1(n)u_n + a_2(n)u_{n-1} + a_3(n)u_{n-2} = 0, \tag{77.9}$$

where

$$a_j(n) \to \alpha_j, \text{ as } n \to \infty.$$

Suppose

$$\alpha_1 x^2 + \alpha_2 x + \alpha_3 = 0 \tag{77.10}$$

has roots distinct in modulus.

Suppose $u_n \neq 0$ for infinitely many n. Then, if neither root is of modulus 1, we have

$$|u_n| = |\ell|^{n(1+o(1))}$$

where ℓ is a root of (77.10).

For a proof, see Milne-Thompson: "The Calculus of Finite Differences", London (1933). The proof actually shows

$$\lim_{n\to\infty} \left| \frac{u_{n+1}}{u_n} \right| = \ell. \ \square$$

Theorem 78: Let $z \notin [0,1]$, then $P_n(z) \neq 0$ for each n. Moreover

$$|P_n(z)| \leq \alpha \left(\frac{1}{z}\right)^{n(1+o(1))}.$$

Proof: The $P_n(x)$ are orthogonal polynomials because if $m < n$, integrate the following by parts n times:

$$\int_0^1 P_m(x)P_n(x)dx = \frac{1}{n!}\int_0^1 x^n(1-x)^n P_m^{(n)}(x)dx = 0, \tag{78.1}$$

because $P_m^{(n)}(x) = 0$, for $n > m$. The orthogonality of P_n implies that all its zeros lie in $[0,1]$.I t is known that orthogonal polynomials have all their zeros in the interval of support for their measure, in this case $[0,1]$. Thus, $P_n(z) \neq 0$ for each n because $z \notin [0,1]$. From the recurrence for P_n (which was homework), the auxiliary limiting polynomial is

$$x^2 + 2(2z-1)x + 1 \tag{78.2}$$

The roots of this polynomial are

$$\frac{-2(2z-1) \pm \sqrt{4(2z-1)^2 - 4}}{2} = -(2z-1) \pm 2\sqrt{z^2 - z}. \tag{78.3}$$

Note that

$$\frac{(1 \pm \sqrt{1-z})^2}{z} = \frac{2 - z \pm 2\sqrt{1-z}}{z},$$

and

$$\alpha\left(\frac{1}{z}\right) = \max \left| \frac{(1 \pm \sqrt{1-z})^2}{z} \right|_{1/z} \tag{78.4}$$

Then by **Poincarè's Lemma**, this theorem follows. $\square$

We can show a lot more:
Theorem 79: For $z \notin [0,1]$,

$$P_n(z) \leq \alpha \left(\frac{1}{z}\right)^n.$$

Proof: By the **Cauchy Integral Formula**:

$$P_n(z) = \frac{1}{2\pi i} \int_{\mathscr{C}} \frac{\omega^n (1-\omega)^n}{(\omega - z)^{n+1}} d\omega. \tag{79.1}$$

Chose $\mathscr{C}$ to be the contour

$$\mathscr{C} := \{\omega \mid |\omega - z| = \sqrt{z^2 - z}\} = \{\omega \mid \omega = z + \sqrt{z^2 - z}e^{i\theta},\ \ 0 \leq \theta \leq 2\pi\}. \tag{79.2}$$

Thus

$$\left.\begin{aligned} d\omega &= \sqrt{z^2 - z} i e^{i\theta} d\theta, \\ \omega - z &= \sqrt{z^2 - z} e^{i\theta}, \\ \omega(1-\omega) &= z - z^2 + (1-2z)\sqrt{z^2 - z} e^{i\theta} - (z^2 - z)e^{2i\theta}. \end{aligned}\right\} \tag{79.3}$$

By (79.1), (79.2), (79.3), we have

$$\begin{aligned} P_n(z) &= \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{(z - z^2) + (1-2z)\sqrt{z^2 - z} e^{i\theta} - (z^2 - z)e^{2i\theta}}{\sqrt{z^2 - z} e^{i\theta}} \right)^n d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left(-\sqrt{z^2 - z}(e^{i\theta} + e^{-i\theta}) + (1-2z) \right)^n d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} (1 - 2z - 2\sqrt{z^2 - z}\cos\theta)^n d\theta. \end{aligned} \tag{79.4}$$

This yields

$$|P_n(z)| \leq \frac{1}{2\pi} \int_0^{2\pi} \alpha \left(\frac{1}{z}\right)^n d\theta = \alpha \left(\frac{1}{z}\right)^n,$$

which is the theorem. □

Theorem 80: (1980, Alladi-Robinson, Crelle) $\log 2$ has irrationality type $\leq 4.622\ldots$.

Proof: We begin by noting that

$$\begin{aligned} \int_0^1 \frac{x^m}{1+x} dx &= \int_0^1 x^m (1 - x + x^2 - x^3 + \cdots) dx \\ &= \frac{x^{m+1}}{m+1} - \frac{x^{m+2}}{m+2} + \frac{x^{m+3}}{m+3} - \cdots \Big|_0^1 \\ &= (-1)^m \left(\log 2 - \sum_{k=1}^{m} \frac{(-1)^{k-1}}{k} \right) \end{aligned} \tag{80.1}$$

Therefore,

$$\int_0^1 \frac{P_n(x)}{1+x} dx = P_n(-1)\log 2 - \frac{a_n}{b_n}, \tag{80.2}$$

where $a_n, b_n \in \mathbb{Z}$ with $b_n \mid \ell_n$, where $\ell_n = \operatorname{lcm} 343\{1, \ldots, n\}$.

Thus,

$$E_n = \ell_n \int_0^1 \frac{P_n(x)}{1+x} dx = q_n \log 2 - p_n, \tag{80.3}$$

where $p_n, q_n \in \mathbb{Z}$,

$$q_n = \ell_n P_n(-1). \tag{80.4}$$

An n-fold integration by parts yields

$$E_n = \ell_n \int_0^1 \frac{x^n (1-x)^n}{(1+x)^{n+1}}. \tag{80.5}$$

Since the integrand is > 0, we have

$$E_n \neq 0. \tag{80.6}$$

Also,

$$\max_{0 \leq x \leq 1} \frac{x(1-x)}{1+x} = (\sqrt{2}-1)^2. \tag{80.7}$$

Thus

$$|E_n| \leq \ell_n (\sqrt{2}-1)^{2n} = \left(e(\sqrt{2}-1)^2\right)^{n(1+o(1))} \quad \to 0 \text{ as } n \to \infty \tag{80.8}$$

because

$$e(\sqrt{2}-1)^2 < 1. \tag{80.9}$$

Thus $\log 2$ is irrational. Observe that $P_n(x)$ is a polynomial with alternating coeffificents. Thus, $P_n(-1) \to \infty$. So, by **Poincarè's Lemma**:

$$P_n(-1) = \alpha(-1)^{n(1+o(1))} = (\sqrt{2}+1)^{2n(1+o(1))}. \tag{80.10}$$

This implies

$$q_{n+1} = q_n^{1+o(1)}, \tag{80.11}$$

with $q_n = (e(\sqrt{2}+1)^2)^{n(1+o(1))}$.

Also since $\|f\| \to \|f\|_\infty$, we have that

$$E_n = \left(e(\sqrt{2}-1)^2\right)^{n(1+o(1))}. \tag{80.12}$$

Therefore

$$\left|\log 2 - \frac{p_n}{q_n}\right| = \frac{\left(e(\sqrt{2}-1)^2\right)^{n(1+o(1))}}{q_n} = \frac{1}{q^{(1+\lambda)(1+o(1))}} \tag{80.13}$$

with

$$\lambda = -\log(e(\sqrt{2}-1)^2)/\log(e(\sqrt{2}+1)^2) \tag{80.14}$$

and

$$1 + \frac{1}{\lambda} = 4.622\ldots. \tag{80.15}$$

Now **Theorem 80** follows from **Theorem 74**. □

Theorem 81: (Lemma on Padé Approximations) Suppose $f(z) \in \mathbb{C}[z]$ such that for all $n \geq n_0$ there exist polynomials $A_n(z)$, $B_n(z)$ of degree $\leq n$ such that

$$A_n(z) + B_n(z)f(z) = z^{2n+1}E_n(z), \tag{81.1}$$

where $E_n(z) \in \mathbb{C}[[z]]$. Suppose further that $E_n(0) \neq 0$ for each $n \geq n_0$, and not both $A_n(0)$ and $B_n(0)$ are 0. Then,

$$A_n(z)B_{n+1}(z) - A_{n+1}(z)B_n(z) = c_n z^{2n+1} \tag{81.2}$$

where $c_n \in \mathbb{C}$ and $c_n \neq 0$.

Proof: Identity for (81.1) is:

$$A_{n+1}(z) + B_{n+1}(z)f(z) = z^{2n+3}E_{n+1}(z). \tag{81.3}$$

Therefore

$$A_n(z)B_{n+1}(z) - A_{n+1}(z)B_n(z) = z^{2n+1}D_n(z). \tag{81.4}$$

where $D_n(z) \in \mathbb{C}[[z]]$, i.e. the right hand side has a zero of order at least $2n+1$. But this is a polynomial of degree $\leq 2n+1$. Thus, the right hand side of (81.4) is

$$c_n z^{2n+1}. \tag{8.15}$$

It remains to show that $c_n \neq 0$.

Suppose $c_n = 0$. Then,

$$A_n(z)B_{n+1}(z) - A_{n+1}(z)B_n(z) \equiv 0 = z^{2n+1}E_n(z)B_{n+1}(z) - z^{2n+1}E_{n+1}(z)B_n(z) \tag{81.6}$$

by (81.1) and (81.3).

Thus,

$$E_n(z)B_{n+1}(z) = z^2 E_{n+1}(z)B_n(z). \tag{81.7}$$

Therefore,

$$z^2 \mid E_n(z)B_{n+1}(z) \tag{81.8}$$

but $E_n(0) \neq 0$. Hence

$$z^2 \mid B_{n+1}(z).$$

Similarly, we will get

$$z^2 \mid E_n(z)A_{n+1}(z)$$

and this yields

$$z^2 \mid A_{n+1}(z) \tag{81.9}$$

because $E_n(z) \neq 0$. Thus, z^2 divides both $A_{n+1}(z)$ and $B_{n+1}(z)$. This implies that

$$A_{n+1}(0) = B_{n+1}(0) = 0,$$

which is a contradiction. Hence, $c_n \neq 0$. $\square$

Theorem 82: For $z \notin [1, \infty)$ let

$$I_n(z) = \int_0^1 \frac{P_n(x)}{1 - zx} dx. \tag{82.0}$$

Then, $I_n(z) \neq 0$ for infinitely many n. Moreover

$$|I_n(z)| = \beta(z)^{n(1+o(1))} \tag{82.1}$$

Proof: Since $P_n(x)$ are orthogonal polynomials, we have that

$$f(x) := \frac{1}{1 - zx} - \sum_{n=0}^{N} \frac{I_n(z) P_n(x)}{\int_0^1 P_n^2(t) dt} \tag{82.2}$$

should satisfy

$$\int_0^1 f(x) P_m(x) dx = 0, \text{ for all } m = 0, 1, 2, \ldots. \tag{82.3}$$

Since the $P_n(x)$ form a basis for $\mathbb{C}[x]$, we conclude that

$$\int_0^1 f(x) P(x) dx = 0, \text{ for all } P(x) \in \mathbb{C}[x]. \tag{82.4}$$

Since the polynomials are dense in the space of continuous functions, we conclude that

$$\int_0^1 f(x) g(x) dx = 0, \text{ for all continuous functions } g \in \mathscr{C}[0, 1].$$

Hence, $f \equiv 0$. But this implies that $1/(1 - zx) \in \mathbb{C}[x]$, which is a contradiction.
Therefore, $I_n(z) \neq 0$ for at least infinitely many $n \in \mathbb{N}$.

Next we establish a recurrence for $I_n(z)$, namely consider

$$nI_n(z) + (2n-1)\left(\frac{2}{z} - 1\right) I_{n-1}(z) + (n-1) I_{n-2}(x) =$$

$$\int_0^1 \frac{nP_n(x) + (2n-1)\left(\frac{2}{z} - 1\right) P_{n-1}(x) - (n-1) P_{n-2}(x)}{1 - zx} dx. \tag{82.5}$$

rom the recurrence

$$nP_n(x) + (n-1) P_{n-2}(x) = -(2n-1)(2x-1) P_{n-1}(x) \tag{82.6}$$

for the Legendre polynomials, we get that

$$nI_n(z) + (2n-1)\left(\frac{2}{z} - 1\right) I_{n-1}(z) + (n-1) I_{n-2}(z) = (2n-1)\frac{2}{z} \int_0^1 P_{n-1}(x) dx = 0 \tag{82.7}$$

because $n - 1 \geq 0$.

Rewrite the recurrence for $I_n(z)$ as

$$I_n(z) + \left(2 - \frac{1}{n}\right)\left(\frac{2}{z} - 1\right) I_{n-1}(z) + \left(1 - \frac{1}{n}\right) I_{n-2}(z) = 0. \tag{82.8}$$

This is a recurrence with almost constant coefficients, with characteristic polynomial

$$x^2 + 2\left(\frac{2}{z} - 1\right) x - 1 = 0 \tag{82.9}$$

whose roots are $\alpha(z)$ and $\beta(z)$. Since $\alpha(z)\beta(z) = 1$ for all z and $\alpha(z) = \beta(z)$ precisely when $z \in [1, \infty)$, we infer that

$$\alpha(z) > 1 > \beta(z). \tag{87.10}$$

Now, by **Poincaré's Theorem**,

$$|I_n(z)| = \ell^{n(1+o(1))} \tag{82.11}$$

with $\ell := \alpha(z)$ or $\ell := \beta(z)$.

Since $\dfrac{1}{1-zx}$ is constant in $[0, 1]$, we know that $I_n(z) \to 0$ as $n \to \infty$. Thus

$$|I_n(z)| = \beta(z)^{n(1+o(1))}. \ \square$$

Theorem 83: (Alladi-Robinson, 1980, Crelle) Let K be either the rationals or an imaginary quadratic field. Let R be the ring of integers in K. Suppose $r, s \in R$ satisfy

$$r/s \notin [1, \infty), \tag{83.1}$$

$$\beta(r/s)|r| \cdot e < 1. \tag{83.2}$$

Then,

$$\left|\log\left(1 - \frac{r}{s}\right) - \frac{a}{b}\right| > \frac{1}{b^{\tau+\epsilon}}, \tag{83.3}$$

for all $a, b \in R$, $|b| > b_0(\epsilon)$, where

$$\tau = 1 + \frac{\log\left(\alpha\left(\dfrac{r}{s}\right) \cdot |r|\right) + 1}{\log\left(\alpha\left(\dfrac{r}{s}\right) \cdot r\right) - 1}. \tag{83.4}$$

Proof: First observe that if $|z| < 1$ then

$$\begin{aligned}\int_0^1 \frac{x^m}{1-zx}dx &= \int_0^1 x^m(1 + zx + z^2x^2 + \cdots)dx \\ &= \sum_{k=0}^{\infty} z^k \int_0^1 x^{m+k}dx \\ &= \sum_{k=0}^{\infty} \frac{z^k}{m+k+1}.\end{aligned} \tag{83.5}$$

Since

$$\log(1-z) = -\sum_{k=0}^{\infty} \frac{z^{k+1}}{k+1}. \tag{83.6}$$

we have

$$\int_0^1 \frac{x^m}{1-zx}dx = -\frac{1}{z^{m+1}}\left(\log(1-z) + \sum_{k=1}^{m} \frac{z^k}{k}\right). \tag{83.7}$$

Note that (83.7) holds for $m = 0$ as well. The left integral in (83.7) is analytic in z for $z \notin [1, \infty)$.

Next, we can define a single-valued branch of $\log(\omega)$ for $\omega \in \mathbb{C} \setminus [0, -\infty)$, i.e., $\log(1-z)$ can be defined in $z \notin [1, \infty)$. Thus

$$\text{(83.7) holds for all } z \in \mathbb{C} \setminus [1, \infty). \tag{83.8}$$

Consequently,

$$I_n(z) := \int_0^1 \frac{P_n(x)}{1-zx}dx = -\frac{1}{z}P_n\left(\frac{1}{z}\right)\log(1-z) + \frac{1}{\ell_n}Q_n\left(\frac{1}{z}\right) \tag{83.9}$$

where $\ell_n = \text{lcm}\{1, 2, 3, \ldots, n\}$, with Q_n satisfying

$$Q_n(x) \in \mathbb{Z}[x], \qquad x \mid Q_n(z), \qquad \deg(Q_n) = n. \tag{83.10}$$

An n-fold integration by parts of I_n yields:

$$I_n(z) = (-z)^n \int_0^1 \frac{x^n(1-x)^n}{(1-zx)^{n+1}} dx. \tag{83.11}$$

From (83.10) and (83.11), we have

$$I_n(z) = (-z)^n \int_0^1 \frac{x^n(1-x)^n}{(1-zx)^{n+1}} dx = -\frac{1}{z} P_n\left(\frac{1}{z}\right) \log(1-z) + \frac{1}{\ell_n} Q_n\left(\frac{1}{z}\right). \tag{83.12}$$

Now define

$$A_n(z) = z^{n+1} Q_n\left(\frac{1}{z}\right), \tag{83.13a}$$

$$B_n(z) = -\ell_n z^n P_n\left(\frac{1}{z}\right), \tag{83.13b}$$

$$E_n(z) = (-1)^n \ell_n \int_0^1 \frac{x^n(1-x)^n}{(1-zx)^{n+1}} dx. \tag{83.14}$$

This yields

$$A_n(z) + B_n(z) \log(1-z) = z^{2n+1} E_n(z). \tag{83.15}$$

Clearly

$$E_n(0) \neq 0, \tag{83.16}$$

and

$$B_n(0) = -\ell_n \binom{2n}{n} \neq 0. \tag{83.17}$$

So, by the lemma on *Padé Approximations*, we have that

$$A_{n+1}\left(\frac{r}{s}\right) B_n\left(\frac{r}{s}\right) \neq A_n\left(\frac{r}{s}\right) B_{n+1}\left(\frac{r}{s}\right), \tag{83.18}$$

with $z = r/s$. We clear the denominators caused by r/s, namely

$$p_n = -s^n A_n\left(\frac{r}{s}\right), \tag{83.19a}$$

$$q_n = s^n B_n\left(\frac{r}{s}\right), \tag{83.19b}$$

where $p_n, q_n \in R$, to get

$$\left| q_n \log\left(1 - \frac{r}{s}\right) - p_n \right| = \left|\frac{r}{s}\right| |r|^n \ell_n \left| I_n\left(\frac{r}{s}\right) \right| \neq 0, \tag{83.20}$$

and

$$|q_n| \leq e \cdot |r| \cdot \alpha \left(\frac{r}{s}\right)^{n(1+o(1))}, \tag{83.21}$$

$$\left| q_n \log\left(1 - \frac{r}{s}\right) - p_n \right| \leq \left(er\beta\left(\frac{r}{s}\right)\right)^{n(1+o(1))} \longrightarrow 0 \text{ as } n \to \infty. \tag{83.22}$$

This proves the theorem. □

Remark: The method used above applies if $1 - zx$ in the denominator is replaced by $(1 - zx)^{\ell/k}$. To do this, we need an estimate on the least common multiple

$$d_n(k, \ell) := \mathrm{lcm}_{m=1,\ldots,n}[km + k - \ell]. \tag{84.1}$$

Corollary 1: $\log 2$ has irrationality type $\leq 4.622\ldots$.

Corollary 2: If $p, q \in \mathbb{Z}^+$ and $\left(1 - \sqrt{1 + \frac{p}{q}}\right)^2 \cdot q \cdot e < 1$, then $\log\left(1 + \frac{p}{q}\right)$ has irrationality type $\leq$:

$$z_{p,q} = 1 + \frac{\log\left(q\left(1 + \sqrt{1 + \frac{p}{q}}\right)^2\right) + 1}{\log\left(q\left(1 + \sqrt{1 + \frac{p}{q}}\right)^2\right) - 1}.$$

Remark: If $q \to \infty$ and $\log(p)/\log(q) \to 0$, then $z_{p,q} \to 2$.

Corollary 3: $\pi/\sqrt{3}$ is irrational of type ≤ 8.3099.

Theorem 84: Let $1 \leq \ell \leq k$, with

$$\gcd(\ell, k) = 1. \tag{84.1}$$

Let

$$f(k) = \frac{k}{\varphi(k)} \sum_{\substack{j=1 \\ (j,k)=1}}^{k} \frac{1}{j}. \tag{84.2}$$

Then

$$d_n(k, \ell) = e^{f(k)n(1+o(1))}. \tag{84.3}$$

Proof: Let $x := kn + k - \ell$ and let p be prime. We choose n sufficiently large that if $\sqrt{x} < p < x$¡ then $\gcd(p, k) = 1$. Let α_p be the largest power of p dividing any one of the numbers $km + k - l$ with $m = 1, \ldots, n$. Clearly

$$\alpha_p \leq \frac{\log(x)}{\log(2)}. \tag{84.4}$$

Note that

$$d_n(k, \ell) = \prod_{p \text{ prime}} p^{\alpha_p} = \prod_{p \leq \sqrt{x}} p^{\alpha_p} \cdot \prod_{\sqrt{x} < p < x} p = \Pi_1 \cdot \Pi_2. \tag{84.5}$$

Now,

$$\Pi_1 \leq \prod_{p \leq \sqrt{x}} p^{\log(x)/\log(2)} = e^{\left\{\sum_{p \leq \sqrt{x}} (\log(p))\right\}(\log(x)/\log(2))} = e^{\theta(x)(\log(x)/\log(2))} = e^{\sqrt{x}\log(x)(1+o(1))}. \tag{84.6}$$

Next

$$\Pi_2 = \prod_{\substack{p > \sqrt{x} \\ p| \text{ some } km + k - \ell \leq x}} p = \prod_{\substack{j=1 \\ (j,k)=1}}^{k} \left(\prod_{\substack{p \equiv j \pmod k \\ p| \text{ some } km + k - \ell}} p \right). \tag{84.7}$$

Claim 84.8:

$p \mid km+k-\ell$ for some m with $km+k-\ell \le x$ and $p \equiv j \pmod k$ if and only if $p \le \dfrac{x}{a_j}$, where a_j is the least positive residue of $-\ell j^{-1} \pmod k$. To justify this claim, observe that

$$\begin{aligned} & p \mid km+k-\ell \text{ for some } m, \text{ with } km+k-\ell \le x \text{ and } p \equiv j \pmod k \\ \iff & \lambda p = km+k-\ell,\ k \in \mathbb{Z},\ m \le n,\ p \equiv j \pmod k \\ \iff & \lambda p \equiv -\ell \pmod k,\ \lambda p \le x,\ k \in \mathbb{Z},\ m \le n,\ p \equiv j \pmod k \\ \iff & \lambda p \equiv -\ell j^{-1} \pmod k,\ \lambda p \le x,\ k \in \mathbb{Z},\ m \le n,\ p \equiv j \pmod k \end{aligned} \tag{84.9}$$

The best way to achieve this is to choose $\lambda = a$, and therefore $p \le x/a_j$ as claimed.

Therefore,

$$\begin{aligned} \Pi_2 &= \prod_{\substack{j=1 \\ (j,k)=1}}^{k} \left(\prod e^{\left\{ \sum\limits_{\substack{p \le x/a_j \\ p \equiv j \pmod k}} \log(p) \right\}} \right) \\ &= e^{\left\{ \sum\limits_{\substack{j=1 \\ (j,k)=1}}^{k} \dfrac{x}{a_j \varphi(k)} \right\}} \end{aligned} \tag{84.10}$$

Note that a_j runs through the full set of reduced residues mod k. Hence the theorem holds. □

Theorem 85: (Alladi-Robinson, Crelle, 1980) Let $\ell, k, r, s \in \mathbb{N}$, with $\ell < k$, and $\gcd(\ell, k) = 1$, $\gcd(r, s) = 1$, and $r < s$. Suppose

$$\left(1 - \sqrt{1 + \frac{r}{s}}\right)^2 s \cdot e^{f(k)} < 1, \tag{85.1}$$

where $f(x) = \dfrac{k}{\varphi(k)} \sum\limits_{\substack{j=1 \\ (j,k)=1}}^{K} \dfrac{1}{j}$. Then, $\left(1 + \dfrac{r}{s}\right)^{-\ell/k}$ is irrational of type

$$\le \tau_{r,s,k} := \frac{\log\left(s\left(1 + \sqrt{1 + \dfrac{r}{s}}\right)^2\right) + f(k)}{\log\left(\dfrac{s}{r^2}\left(1 + \sqrt{1 + \dfrac{r}{s}}\right)^2\right) - f(k)} + 1. \tag{85.2}$$

Proof: Using the substitution $u := 1 - zx$. we get that

$$\int_0^1 \frac{x^m}{(1-zx)^{\ell/k}} dx = -\frac{1}{z} \cdot \frac{1}{z^m} \int_1^{1-z} (1-u)^m u^{-(\ell/k)} du =$$

$$= \frac{1}{z^{m+1}} \int_{1-z}^{1} \left(\sum_{j=0}^{m} (-1)^j \binom{m}{j} u^{j-(\ell/k)} \right) du = \frac{1}{z^{m+1}} \sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{1-(1-z)^{j+1-(\ell/k)}}{j+1-(\ell/k)}.$$

85.3

(85.3)

If we define $a_m(n)$ by

$$P_n(x) = \sum_{m=0}^{n} a_m(n)x^m, \tag{85.4}$$

then

$$d_n(k,\ell)\cdot z^{n+1}\cdot\int_0^1 \frac{P_n(x)}{(1-zx)^{\ell/k}}dx = A_n(z) + B_n(z)(1-z)^{1-(\ell/k)}, \tag{85.5}$$

where

$$A_n(z) = d_n(k,\ell)\cdot\sum_{m=0}^{n} a_m(n)z^{n-m}\sum_{j=0}^{m}(-1)^j\binom{m}{k}\frac{1}{j+1-(\ell/k)}, \tag{85.6}$$

and

$$B_n(z) = d_n(k,\ell)\cdot\sum_{m=0}^{n} a_m(n)z^{n-m}\sum_{j=0}^{m}(-1)^j\binom{m}{k}\frac{1}{j+1-(\ell/k)}. \tag{85.7}$$

So, now we have that

$$A_n(z), B_n(z)\in\mathbb{Z}[z] \qquad \text{and} \qquad \deg A_n = \deg B_n = n. \tag{85.8}$$

An n-fold integration by parts yields

$$\int_0^1 \frac{P_n(x)}{(1-zx)^{\ell/k}}dx = \frac{(-z)^n}{n!}\cdot\prod_{j=1}^{n}(j+(\ell/k)-1)\int_0^1\frac{x^n(1-x)^n}{(1-zx)^{n+(\ell/k)}}dx. \tag{85.9}$$

Now let

$$E_n(z) = (-1)^n d_n(k,\ell)\frac{1}{n!}\prod_{j=1}^{n}(j+(\ell/k)-1)\int_0^1\frac{x^n(1-x)^n}{(1-zx)^{n+(\ell/k)}}dx. \tag{85.10}$$

Thus

$$A_n(z) + B_n(z)(1-z)^{1-(\ell/k)} = z^{2n+1}E_n(z). \tag{85.11}$$

Also,

$$E_n(0) = 0 \tag{85.12}$$

because the integrand is positive (if $0<z<1$), and we're going to choose $z = r/s$.

Moreover,

$$A_n(0) = d_n(k,\ell)a_m(n)\int_0^1\frac{(1-y)^n}{y^{\ell/k}}dy > 0. \tag{85.13}$$

Thus by **Theorem 81** (Padé Approximation Lemma), we have that

$$A_n\left(-\frac{r}{s}\right)B_{n+1}\left(-\frac{r}{s}\right) \neq A_{n+1}\left(-\frac{r}{s}\right)B_n\left(-\frac{r}{s}\right). \tag{85.14}$$

Next define

$$p_n := -s^n A_n\left(-\frac{r}{s}\right) \qquad \text{and} \qquad q_n := s^n B_n\left(-\frac{r}{s}\right). \tag{85.15}$$

Thus,

$$q_n\left(1+\frac{r}{s}\right)^{1-(\ell/k)} - p_n = s^n\left(\frac{r}{s}\right)^{2n+1}d_n(k,\ell)\frac{(-1)^{n+1}}{n!}\prod_{j=1}^{m}(j+(\ell/k)-1)\int_0^1\frac{x^n(1-x)^n}{\left(1+\frac{r}{s}\ell\right)^{n+(\ell/k)}}dx. \tag{85.16}$$

Note that

$$0 < \frac{1}{n!}\prod_{j=1}^{n}(j+(\ell/k)-1) < 1. \tag{85.17}$$

Also,

$$0 < \int_0^1 \frac{x^n(1-x)^n}{\left(1+\frac{r}{s}\ell\right)^{n+(\ell/k)}}dx < \int_0^1 \frac{x^n(1-x)^n}{\left(1+\frac{r}{s}x\right)^{n}}dx \le \left(\frac{s}{r}\right)^{2n}\left(1-\sqrt{1+\frac{r}{s}}\right)^{2n}, \tag{85.18}$$

because

$$\max_{0\le x\le 1}\frac{x(1-x)}{1+\frac{r}{s}x} = \left(\frac{s}{r}\right)^2\left(1-\sqrt{1+\frac{r}{s}}\right)^2. \tag{85.19}$$

Hence if (85.1) is satisfies, the expansion in (85.16) is nonzero and approaches zero as n approaches infinity. Also,

$$0 \le \left|q_n\left(1+\frac{r}{s}\right)^{1-(\ell/k)} - p_n\right| \le \left(se^{f(k)}\left(1-\sqrt{1+\frac{r}{s}}\right)^2\right)^{n(1+o(1))}. \tag{85.20}$$

We need an upper bound for q_n in order to get an irrationality measure.

Note that

$$\left|\sum_{j=0}^{m}(-1)^j\binom{m}{j}\frac{\left(1+\frac{r}{s}\right)^k}{j+1-(\ell/k)}\right| = \left(1+\frac{r}{s}\right)^{(\ell/k)-1}\int_0^1\frac{(1-u)^m}{u^{\ell/k}}du \le \int_0^2 u^{-(\ell/k)}du \le 2k. \tag{85.21}$$

Also, $\operatorname{sgn} a_m(n) = (-1)^m$. Therefore,

$$\left|B_n\left(-\frac{r}{s}\right)\right| \le 2k\left(\frac{r}{s}\right)^n P_n\left(-\frac{r}{s}\right)d_n(k,\ell), \tag{85.22}$$

and

$$q_n \le \left(r\alpha\left(-\frac{r}{s}\right)e^{f(x)}\right)^{n(1+o(1))}. \tag{85.23}$$

Now apply the **Lemma** on Padé Approximations with

$$Q = r\alpha\left(-\frac{r}{s}\right)e^{f(k)}, \qquad E^{-1} = se^{f(k)}\left(1-\sqrt{1+\frac{r}{s}}\right)^2.$$

The theorem follows. □

Remark: If $\frac{r}{s}$ is small enough, i.e. $\frac{r^2}{s}$ is small enough, then (85.1) will hold. Furthermore, if $\frac{\log(r)}{\log(s)}$ is sufficiently close to zero, we get that $\tau_{r,s,k} < 3$.

Corollary: $\sqrt[3]{17}$ is irrational of type $\le 2.5763\cdots$.

Proof: Take $r = 1$, $s = 5831$. Note that $s+1 = s+r = 5832 = 18^3$, and $5831 = 17\cdot 7^3$. So, $1+\frac{r}{s} = \frac{18^3}{17\cdot 7^3}$. Thus, $\sqrt[3]{1+\frac{r}{s}} = \frac{18}{7}\cdot\frac{1}{\sqrt[3]{17}}$. We apply the theorem to $\sqrt[3]{1+\frac{r}{s}}$ to get an irrationality measure < 3, and conclude that $\sqrt[3]{17}$ has irrationality measure < 3. □

Remark: A corollary to this is that $x^3 - 17y^3 = M$ has only a finite number of solutions.

Theorem: (Baker) $\sqrt[3]{2}$ has irrationality type < 3. He proved this by considering $128 = 2 \cdot 4^3$ and $125 = 5^3$. So, $2 \cdot 4^3 \sim 5^3$. Now,

$$\sqrt[3]{\frac{128}{125}} = \frac{4}{5}\sqrt[3]{2}, \text{ etc} \ldots .$$

His theorem was deeper and can use this to prove the result. Our previous theorem cannot do this.

Remark: The corollary to this is that $x^3 - 2y^3 = M$ has only a finite number of solutions.

Theorem 88: (We are skipping a few theorems that we'll go back to.) Let D be not a k^{th} power. Then,

$$x^k - Dy^k = M$$

has only a finite number of solutions.

Proof: Suppose not. Then the equation above has infinitely many solutions. Pick an arbitrary large solution $(x, y) = (x_0, y_0)$. Then, $x_0^k \sim Dy_0^k$, and so $\frac{x_0}{y_0} \sim \sqrt[k]{D}$, and thus $\frac{Dy_0^k}{x_0^k} = 1 + \frac{r}{s}$ where $\frac{r}{s}$ is arbitrarily small. By **Theorem 85**, the irrationality measure is $< k$. Hence there are actually a finite number of solutions. This is a contradiction. □

11.4 $\zeta(2)$

Theorem 89: $\zeta(2)$ is irrational, and of type ≤ 11.85.

Proof: (due to Beukers, originally of Ápery, details by Alladi)

Lemma L.89: Let r, s be integers ≥ 0, and let

$$a_{r,s} = \int_0^1 \int_0^1 \frac{x^r y^s}{1-xy} dxdy. \tag{89.1}$$

Then,

(1) $a_{r,s}$ is a rational with denominator dividing ℓ_r^2 if $r > s$. Recall that $\ell_r := \text{lcm}\{1, \ldots, r\}$.

(2) $a_{r,r} = \zeta(2) - \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{r^2}\right)$.

Proof: Writing:

$$\int_0^1 \int_0^1 x^r y^s (1 + xy + x^2y^2 + \cdots) dxdy$$

we see that this is equal to

$$\int_0^1 \int_0^1 (x^r y^s + x^{r+1} y^{s+1} + x^{r+2} y^{s+2} + \cdots) dxdy.$$

$$= \frac{1}{(r+1)(s+1)} + \frac{1}{(r+2)(s+2)} + \cdots. \tag{89.2}$$

If $r > s$, then we may way rewrite (89.2) as

$$a_{r,s} = \frac{1}{r-s} \left\{ \left(\frac{1}{s+1} - \frac{1}{r+s} \right) + \left(\frac{1}{s+2} - \frac{1}{r+2} \right) + \cdots \right\}.$$

But, this telescopes, and so

$$a_{r,s} = \frac{1}{r-s} \left\{ \frac{1}{s+1} + \frac{1}{s+2} + \cdots + \frac{1}{r} \right\}, \tag{89.3}$$

which is clearly a rational with denominator divising ℓ_r^2.

When $r = s$, we have that

$$a_{r,r} = \frac{1}{(r+1)^2} + \frac{1}{(r+2)^2} + \frac{1}{(r+3)^2} + \cdots = \zeta(2) - \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{r^2}\right). \tag{89.4}$$

□

Now, consider the expression

$$E_n = \int_0^1 \int_0^1 \frac{P_n(x)(1-y)^n}{1-xy} dxdy. \tag{89.5}$$

Note that $P_n(x) \in \mathbb{Z}[x]$ and $\deg P_n = n$. Also, $(1-y)^n \in \mathbb{Z}[y]$ and $\deg(1-y)^n = n$. Lastly, the above expression is a integer linear combination of $a_{r,s}$ for $0 \leq r, s \leq n$. Hence, ℓ_n^2 clears the denominators. Thus by **Lemma L.89**,

$$E_n = q_n \zeta(2) - p_n, \text{ for some } p_n q_n \in \mathbb{Z}. \tag{89.6}$$

Next, and n-fold integration by parts with respect to x yields

$$E_n = \ell_n^2 \int_0^1 \int_0^1 \frac{x^n (1-x)^n y^n (1-y)^n}{(1-xy)^{n+1}} dxdy. \tag{89.7}$$

Note that the integrand is positive. Hence

$$E_n > 0. \tag{89.8}$$

Next we compute

$$\max_{\substack{0\le x\le 1\\ 0\le y\le 1}} f(x,y) =: M_f. \tag{89.9}$$

where

$$f(x,y) := \frac{x(1-x)y(1-y)}{1-xy}. \tag{89.10}$$

Clearly, the max is attained when

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0. \tag{89.11}$$

Observe that

$$\frac{\partial x}{\partial y} = \frac{(1-xy)(1-y)(1-2x) - x(1-x)y(1-y)(-y)}{(1-xy)^2}. \tag{89.12}$$

Thus with some cancellation,

$$\frac{\partial f}{\partial y} = 0 \Longrightarrow (1-xy)(1-2y) + xy(1-y) = 0. \tag{89.13}$$

Comparing (89.12) and (89.13), we get

$$x = y, \tag{89.14}$$

i.e., we want

$$(1-x^2)(1-2x) + x^2(1-x) = 0.$$

This is true if and only if

$$x^3 - 2x + 1 = (x-1)(x^2+x-1) = 0. \tag{89.15}$$

Hence, $x = 1$ and $\beta := \dfrac{\sqrt{5}-1}{2}$ are roots.

So, the maximum is at β:

$$M_f = f(\beta) = \frac{\beta(1-\beta)\beta(1-\beta)}{1-\beta^2}. \tag{89.16}$$

But, $1-\beta = \beta^2$, and so

$$f(\beta) = \frac{\beta^6}{\beta} = \beta^5 = \left(\frac{\sqrt{5}-1}{2}\right)^5. \tag{89.17}$$

Now use $\|g\|_n \to \|g\|_\infty$ to get that

$$E_n = \left(e^2\beta^5\right)^{n(1+o(1))}. \tag{89.17}$$

Observe that

$$0 < e^2\beta^5 < 1. \tag{89.18}$$

Hence $E_n \to 0$ as $n \to \infty$. Thus $\zeta(2) \notin \mathbb{Q}$. To get an irrationality type for $\zeta(2)$, we need an estimate on the size of q_n. For this, we write

$$P_n(x) = \sum_{m=0}^{n} a_m(n) x^m. \tag{89.19}$$

Then

$$q_n = \ell_n^2 \sum_{m=0}^{n} |a_m(n)| \binom{n}{m}. \tag{89.20}$$

So,

$$\begin{aligned} P_n(x) &= \frac{1}{n!}\frac{d^n x^n(1-x^n)}{dx^n} \\ &= \frac{1}{n!}\frac{d^n}{dx^n}\left(\sum_{m=0}^{n}\binom{n}{m}(-1)^m x^{n+m}\right) \\ &= \sum_{m=0}^{\infty}(-1)^m x^m \binom{n}{m}\frac{(n+m)!}{m!n!}. \end{aligned} \tag{89.21}$$

Thus,

$$a_m(n) = (-1)^m \binom{n}{m}\binom{n+m}{m}, \tag{89.22}$$

yielding

$$q_n = \ell_n^2 \sum_{m=0}^{n} \binom{n}{m}^2 \binom{n+m}{m}. \tag{89.23}$$

This expression is due to Apery.

How do we estimate q_n from this? What we need is an asymptotic estimate for $\log q_n$. To get this, it suffices to estimate

$$\max_{0\le m\le n} \binom{n}{m}^2 \binom{n+m}{m}.$$

To this end. we use the Weak Stirling Formula:

$$\log(n!) \sim n\log n - n. \tag{89.24}$$

Set $m := \lambda n$, for $0 < \lambda < 1$. Now,

$$\binom{n}{m}^2 \binom{n+m}{m} = \frac{n!(n+m)!}{(m!)^3((n-m)!)^2}$$

So, we get

$$n\log(n) - n + (m+n)\log(m+n) - n - m \sim e.$$

Now,

$$\begin{aligned} e^{3m\log(m)-3m+2(n-m)\log(n-m)-2(n-m)} &= e^{n\log(n)+(n+m)\log(n+m)-3m\log(m)-2(n-m)\log(n-m)} \\ &= e^{n\log(n)+n(1+\lambda)(\log(n)+\log(1+\lambda))-3n\lambda(\log(n)+\log(\lambda))+2n(1-\lambda)(\log(n)+\log(n-\lambda))} \end{aligned} \tag{89.25}$$

Note that the coefficient of $n\log(n)$ is

$$1 + (1+\lambda) - 3y - 2(1-\lambda) = 0. \tag{89.26}$$

The coefficient of n is

$$g(\lambda) = (1+\lambda)\log(1+\lambda) - 3\lambda\log(\lambda) - 2(1-\lambda)\log(1-\lambda). \tag{89.27}$$

Thus with $m = \lambda n$, we have that

$$\binom{n}{m}^2 \binom{n+m}{m} = e^{nf(\lambda)(1+o(1))}. \tag{89.28}$$

Observe that

$$f'(\lambda) = 1 + \log(1+\lambda) - 3 - 3\log(\lambda) + 2 + 2\log(1-\lambda) = 0.$$

Therefore

$$\frac{(1+\lambda)(1-\lambda)^2}{\lambda^3} = 1,$$

and hence

$$(1+\lambda)(1-\lambda)^2 = \lambda^3.$$

Thus,

$$1 - \lambda - \lambda^2 = 0,$$

which has the above β as a root.

It turns out that $f(\beta) = 5\log(\alpha)$, where $\alpha = \dfrac{1+\sqrt{5}}{2}$. Hence

$$q_n = (e^2\alpha^5)^{n(1+o(1))}. \tag{89.29}$$

By Alladi's earlier Lemma, this yields that the irrationality measure is $5\log(\alpha) \approx 11.85$.

11.5 $\zeta(3)$

Theorem 92: (Apéry) $\zeta(3)$ is irrational of type ≤ 13.8.

Proof: (Beukers) We begin with the observation:

$$\frac{d}{d\sigma}\left(\sum_{n=1}^{\infty}\frac{1}{(n+\sigma)^s}\right)\Bigg|_{\sigma=0} = \sum_{n=1}^{\infty}\frac{-s}{(n+\sigma)^{s+1}}\Bigg|_{\sigma=0} = -s\zeta(s+1). \tag{92.1}$$

From this view point, $\zeta(s+1)$ comes out of $\zeta(s)$ via a differentiation process.

Lemma (L-92):

For integers $r, s \geq 0$, define

$$I_{r,s} := \int_0^1\int_0^1\int_0^1 \frac{x^r y^s}{1-(1-xy)z}\,dx\,dy\,dz. \tag{92.2}$$

Then

(1) $I_{r,s}$ is a rational with denominator dividing ℓ_r^3 if $r \geq s$. (92.3)

(2) $I_{r,r} = 2\left(\zeta(3) - 1 - \frac{1}{2^3} - \cdots - \frac{1}{r^3}\right)$. (92.4)

Proof of Lemma:

Note that

$$\int_0^1 \frac{1}{1-(1-xy)z}dx = \frac{\log(1-(1-xy)z)}{1-xy}\Bigg|_0^1 = -\frac{\log(xy)}{1-xy}$$

Therefore,

$$I_{r,s} = -\int_0^1\int_0^1 \frac{\log(xy)\cdot x^r y^s}{1-xy}\,dx\,dy. \tag{92.5}$$

When $r \geq s$, we know that

$$a_{r,s} = \int_0^1\int_0^1 \frac{x^r y^s}{1-xy}\,dx\,dy = \frac{1}{r-s}\left\{\left(\frac{1}{s+1}-\frac{1}{r+1}\right)+\left(\frac{1}{s+2}-\frac{1}{r+s}\right)+\cdots\right\}. \tag{92.6}$$

We note that for representation (92.6) to be valid, it is not necessary for r, s to be integres. However, if $r, s \in \mathbb{Z}$, then telescoping takes place. Thus, from (92.6):

$$-\frac{d}{dr}a_{r+\sigma,\,s+\sigma}\big|_{\sigma=0} = \frac{1}{r-s}\left\{\left(\frac{1}{(s+1)^2}-\frac{1}{(r+1)^2}\right)+\left(\frac{1}{(s+2)^2}+\frac{1}{(r+2)^2}\right)+\cdots\right\}. \tag{92.7}$$

If r, s are integers, then this telescopes to

$$\frac{1}{r-s}\left\{\frac{1}{(s+1)^2}+\frac{1}{(s+2)^2}+\cdots+\frac{1}{r^2}\right\}, \tag{92.8}$$

which is a rational with denominator dividing ℓ_r^3. Thus (1) is proved.

To establish (2), consider

$$a_{r,r} = \frac{1}{(r+1)^2}+\frac{1}{(r+2)^2}+\cdots.$$

Therefore,

$$\frac{d}{d\sigma}a_{r+\sigma,\,s+\sigma}\big|_{\sigma=0} = -2\left(\frac{1}{(r+1)^3}+\frac{1}{(r+2)^3}+\cdots\right) = -2\left(\zeta(e)-\left(1+\frac{1}{2^3}+\cdots+\frac{1}{r^3}\right)\right), \tag{92.9}$$

which proved (92.4). □

Analogous to (89.5), we now consider

$$E_n = \ell_n^3 \int_0^1 \int_0^1 \int_0^1 \frac{P_n(x)P_n(y)}{1-(1-xy)z}\, dx\, dy\, dx. \tag{92.10}$$

From **Lemma L-92**, it follows that

$$E_n = q_n\zeta(3) - p_n, \quad p_n, q_n \in \mathbb{Z}. \tag{92.11}$$

From our earlier discussion ,we see that in (92.11)

$$q_n = \ell_n^3 \sum_{m=0}^{n} a_m(n)^2 = \ell_n^3 \sum_{m=0}^{n} \binom{n}{m}^2 \binom{n+m}{m}^2. \tag{92.12}$$

Recall from the proof of log(2) that since $a_m(n)$ alternates in sign:

$$P_n(-1) = \sum_{m=0}^{m} \binom{n}{m}\binom{n+m}{m} = (\sqrt{2}+1)^{2n(1+o(1))}. \tag{92.13}$$

Thus it follows from (92.13) that

$$\max_{0\le m\le n} \binom{n}{m}\binom{n+m}{m} = (\sqrt{2}+1)^{2n(1+o(1))}. \tag{92.14}$$

So from (92.14), we get

$$\max_{0\le m\le n} \binom{n}{m}^2\binom{n+m}{m}^2 = (\sqrt{2}+1)^{4n(1+o(1))}. \tag{92.15}$$

Consequently (92.12) and (92.15) yield

$$q_n = \left\{e^3(\sqrt{2}+1)^4\right\}^{n(1+o(1))}. \tag{92.16}$$

Note that the expression for q_n in (92.12) in Apéry's version of the proof.

Our real task is to evaluate E_n and show that it goes to zero. We integrate E_n by parts with respect to x, n times, to get

$$E_n = \ell_n^3(-1)^n \int_0^1 \int_0^1 \int_0^1 \frac{x^n(1-x)^n P_n(y) y^n z^n}{(1-(1-xy)z)^{n+1}}\, dx\, dy\, dz. \tag{92.17}$$

At this stage, we use the substitution

$$w = \frac{1-z}{1-(1-xy)z}, \tag{92.18}$$

which is equivalent to

$$z + w - 1 = (1-xy)wz, \tag{92.19}$$

which is symmetric in w and z (and also symmetric in x and y, but we don't need this).

So, we can invert w and z in (92.18) to yield:

$$z = \frac{1-w}{1-(1-xy)w}. \tag{92.20}$$

This substitution has more miraculous properties:

$$dz = \frac{-(1-(1-xy)w-(1-w)(1-xy)}{(1-(1-xy)w)^2}\,dw = \frac{-xy}{(1-(1-xy)w)^2}\,dw. \tag{92.21}$$

and also

$$1-(1-xy)z = \frac{xy}{1-(1-xy)w}. \tag{92.22}$$

By combining (92.21) and (92.22), we arrive at

$$\frac{1}{1-(1-xy)z}\,dz = -\frac{1}{(1-(1-xy)w}\,dz. \tag{92.23}$$

In addition,

$$1-w = \frac{xyz}{1-(1-xy)z}. \tag{92.24}$$

Why is this miraculous? We can make the following substitution in (92.17) using (92.23) and (92.24):

$$E_n = -\ell_n^3(-1)^n \int_0^1\int_0^1\int_0^1 \frac{P_n(x)P_n(y)(1-w)^n}{1-(1-xy)w}\,dx\,dy\,dw. \tag{92.25}$$

The triple integral in (92.25) is ideally set up for integration by parts n times with respect to y. This yields:

$$E_n = -\ell_n^3 \int_0^1\int_0^1\int_0^1 \frac{(x^n(1-x)^n y^n(1-y)^n w^n(1-w)^n}{(1-(1-xy)w)^{n+1}}\,dx\,dy\,dw. \tag{92.26}$$

Now set

$$\gamma = \max_{\substack{0\le x\le 1\\ 0\le y\le 1\\ 0\le z\le 1}} \frac{x(1-x)y(1-y)w(1-w)}{1-(1-xy)w} \tag{92.27}$$

to get

$$E_n = (e^3\gamma)^{n(1+o(1))} \tag{92.28}$$

because

$$\|f\|_n \to \|f\|_\infty.$$

The integrand in (92.26) is positive. Hence

$$E_n \neq 0. \tag{92.28}$$

Also (see homework)

$$\gamma = (\sqrt{2}-1)^4. \tag{92.30}$$

and (luckily)

$$0 < e^3\gamma < 1. \tag{92.31}$$

Hence $E_n \to 0$ as $n \to \infty$ and thus $\zeta(3)$ is irrational. To get the irrationality measure, apply our lemma with

$$Q = (\sqrt{2}+1)^4$$

$$E^{-1} = e^{-3}(\sqrt{2}+1)^4. \ \square$$

11.6 Transcendence of e

Theorem 93: (Hermite, 1873) e is transcendental.

Proof: (as in *Transcendental Number Theory*, Baker) We begin by noting that if f is a polynomial (of degree m), then for any $z \in \mathbb{C}$, the integral

$$I(z) = \int_0^z e^{z-w} f(w) dw \tag{93.1}$$

can be evaluated by repeated integration by parts.

Since e^{z-w} and $f(w)$ are both entire, the integral can be evaluated along any path connecting 0 to z - we may take the straight line connecting 0 and z. As a first step:

$$\begin{aligned} I(z) = & -e^z f(w)|_0^z + \int_0^z e^{z-w} f'(w) dw \\ & = e^z f(0) - f(z) - e^{z-w} f'(w)\big|_0^z + \int_0^z e^{z-w} f''(w) dw \\ & = e^z (f(0) + f'(0)) - f'(z) - f(z) + \int_0^z e^{z-w} f''(w) dw \end{aligned} \tag{93.2}$$

So, by iteration of (93.2), we get that

$$I(z) = e^z \sum_{k=0}^{m} f^{(k)}(0) - \sum_{j=0}^{m} f^{(j)}(z). \tag{93.3}$$

Now we define $\widehat{f}(w)$ to be the polynomial obtained from w be replacing all the coefficients by their absolute values. Then, it follows (by a crude estimate) that

$$|I(z)| \leq e^{|z|} \widehat{f}(|z|)|z|. \tag{93.4}$$

Suppose now that e were algebraic, and let

$$a_0 + a_1 e + a_2 e^2 + \cdots + a_n e^n = 0, \tag{93.5}$$

i.e., $g(x) := a_0 + a_1 x + \cdots + a_n x^n$ is the minimal polynomial of e, for $a_i \in \mathbb{Z}$.

We will now consider estimates for

$$J = a_0 I(0) + a_1 I(1) + \cdots + a_n I(n), \tag{93.6}$$

where

$$f(x) = x^{p-1}(x-1)^p (x-2)^p \cdots (x_n)^p, \tag{93.7}$$

and p is a large prime to be specified later.

Clearly (93.7) implies that

$$\left.\begin{array}{ll} f^{(j)}(k) = 0, & \text{if } j < p \text{ and } 0 < k \leq n, \\ f^{(j)}(k) = 0, & \text{if } j < p-1 \text{ and } 0 \leq k \leq n. \end{array}\right\} \tag{93.8}$$

On the other hand,

$$f^{(j)}(k) \equiv 0 \pmod{p!}, \text{ for } j \geq p.$$

But then,

$$\left.\begin{array}{ll} f^{(p-1)}(0) \equiv 0 & (\text{mod } (p-1)!) \\ f^{(p-1)}(0) \not\equiv 0 & (\text{mod } p!) \end{array}\right\} \tag{93.9}$$

if p is chosen large enough.

Observe that (93.3) and (93.6) imply that

$$J = \left(\sum f^{(k)}(0)\right)(a_0 + a_1 e + \cdots + a_n e^n) - \sum_{k=0}^{n}\sum_{j=0}^{m} a_k f^{(j)}(k) = -\sum_{k=0}^{n}\sum_{j=0}^{m} a_k f^{(j)}(k). \tag{93.10}$$

Note that $J \in \mathbb{Z}$. So from (93.9) and (93.10), we see that

$$\left.\begin{array}{ll} J \equiv 0 & (\text{mod } (p-1)!) \\ J \not\equiv 0 & (\text{mod } p!) \end{array}\right\} \tag{93.11}$$

Therefore,

$$|J| \geq (p-1)!. \tag{93.12}$$

To get an upper bound for J first note that

$$\widehat{f}(k) \leq n^{p-1}(2n)^{np} \leq (2n)^m \tag{93.13}$$

because $m = np + p - 1$ and because (93.7) implies that

$$|\widehat{f}(x)| \leq |x|^{p-1}(|x|+1)^p(|x|+2)^p \cdots (|x|+n)^p. \tag{93.14}$$

Thus, by (93.4) and (93.6), we have that

$$|J| \leq |a_0| e \widehat{f}(0) + |a_1| e^2 \widehat{f}(1) + \cdots + |a_n| e^n \widehat{f}(n) \leq C^p, \tag{93.15}$$

where

$$C := \max_{0 \leq i \leq n} |a_i| n e^n (2n)^n. \tag{93.16}$$

However, the lower bound (93.12) and the upper bound (93.16) are incompatible if p is large. This is a contradiction, and therefore e is transcendental. □

Corollary:

Let $r/s \in \mathbb{Q}$ with $r/s \neq 0$. Then $e^{r/s}$ is transcendental, hence irrational.

This leads to the question of whether e^α is transcendental for $\alpha \neq 0$ algebraic. This was proved by **Lindemann** from which the transcendence of π follows.

11.7 Transcendence of π

Theorem 94: (Lindemann (1882)) π is transcendental.

Proof: (as in Baker) Suppose π is algebraic. Then, since i is algebraic, so is $i\pi =: \vartheta$. Let ϑ be algebraic of degree d, and let ℓ be the leading coefficient of the minimal polynomial of ϑ. Let

$$\theta_1 = \{\vartheta_1 \cdots \vartheta_d\}$$

be the full set of roots of this minimal polynomial - they are the set of conjugates of ϑ. Consider next the product

$$\left(1 + e^{\vartheta_1}\right)\left(1 + e^{\vartheta_2}\right)\cdots\left(1 + e^{\vartheta_d}\right) = 0, \tag{94.1}$$

since $e^{\vartheta_1} = e^{i\pi} = -1$.

The product in (94.1) is a sum of expressions of the type e^{Θ}, for

$$\Theta = \sum \epsilon_i \vartheta_i. \tag{94.2}$$

where $\epsilon_i \in \{0, 1\}$. Out of these, let precisely n of them be non-zero, and we call these $\{\alpha_1, \ldots, \alpha_n\}$. Thus,

$$\underbrace{2^d - n}_{=:\, q} + \alpha_1 + \cdots + \alpha_n = 0. \tag{94.3}$$

Analogous to the previous proof, we now consider

$$J = I(\alpha_1) + \cdots + I(\alpha_n), \tag{94.4}$$

with

$$I(z) := \int_0^z e^{z-w} f(w) dw,$$

where

$$f(x) := \ell^{np} x^{p-1} (x - \alpha_1)^p \cdots (x - \alpha_n)^p. \tag{94.5}$$

As before, p is a prime to be chosen sufficiently large later.

–End Review from Last Class–

By the same reasonging underlying (93.2), (93.3), (93.8), we get:

$$J = (e^{\alpha_1} + \cdots + e^{\alpha_n}) \sum_{i=0}^{m} f^{(j)}(0) - \sum_{k=1}^{n} \sum_{j=0}^{m} f^{(j)}(\alpha_k) = -q \sum_{j=0}^{m} f^{(j)}(0) - \sum_{k=1}^{n} \sum_{j=0}^{m} f^{(j)}(\alpha_k). \tag{94.6}$$

where $m = \deg f = np + p - 1$.

Note that in (94.6), the expressions involve symmetric functions in the α_k multiplied by ℓ^{np} (or we may treat them as symmetric functions of the $\ell\alpha_k$) and therefore will be integer values. Hence, $J \in \mathbb{Z}$. Observe that

$$f^{(j)}(\alpha_k) = 0, \text{ for all } k \text{ and for } j < p. \tag{94.7}$$

Additionally,

$$f^{(j)}(\alpha_k) \text{ and } f^{(j)}(0) \quad \equiv 0 \pmod{p}! \text{ if } j \geq p.$$

So, the only case we need to consider is $f^{(p-1)}(0)$:

$$\begin{aligned} f^{(p-1)}(0) &= (p-1)!(-\ell)^{np}(\alpha_1 \cdots \alpha_n)^p \\ &\equiv 0 \pmod{(p-1)!} \\ &\not\equiv \pmod{p!} \end{aligned} \tag{94.8}$$

if p is large enough.

Hence J is a nonzero integral multiple of $(p-1)!$, which implies

$$|J| \geq (p-1)!. \tag{94.9}$$

On the other hand

$$|J| \leq |\alpha_1| e^{|\alpha_1|} \widehat{f}(|\alpha_1|) + \cdots + |\alpha_n| e^{|\alpha_n|} \widehat{f}(|\alpha_n|) \leq C^p. \tag{94.10}$$

(Note: C depends on $\alpha_1, \ldots, \alpha_n$ and n, but is independent of p.) Estimates (94.9) and (94.10) are incompatible if p is large. Hence, the transcendence of π follows. □

Remarks: Lindemann actually proved:

Theorem L1: If α is algebraic and non-zero, then e^α is transcendental.

This has an immediate consequence that $i\pi$ (and hence π) is transcendental, because $e^{i\pi} = -1$.

The transcendence of π has the immediate consequence that it is impossible to square the circle, i.e. to construct a square equal in area to a given circle using only a ruler and compass.

Another consequence is that is α is algebraic and nonzero, then $\sin(\alpha)$, $\cos(\alpha)$ and $\tan(\alpha)$ are transcendental (we have previously shown that is α nonzero rational, then the values of the these functions must be irrational).

The real theorem of Lindemann was:

Theorem L2:

For distinct algebraic numbers $\alpha_1, \ldots, \alpha_n$ and any nonzero algebraic $\beta_1, \ldots, b_n$, then we have

$$\beta_1 e^{\alpha_1} + \cdots + \beta_n e^{\alpha_n} \neq 0,$$

i.e., $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are algebraically independent.

Theorem L1 follows immediately from this, setting one of $\alpha_i = 0$ so $\beta_i e^{\alpha_i} = \beta_i$ which is algebraic.

11.8 $\zeta(2\pi)$

Theorem 95: At even positive integers $2k$, the value $\zeta(2k)$ of the Riemann zeta function is a rational (depending on k) multiple of π^{2k}, i.e.

$$\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = \frac{p_k}{q_k}\pi^{2k}, \text{ for } k = 1, 2, \ldots.$$

An immediate corollary to this is that all values $\zeta(2k)$ are irrational because π is transcendental.

Lemma 95-L: Let $g(x) \in C^1[-\pi, \pi]$, i.e., $g(x)$ is differentiable on $[-\pi, \pi]$ and $g'(x)$ is continuous. Let the Fourier coefficients of g be defined by

$$\begin{aligned} a_0 &:= \frac{1}{2\pi}\int_{-\pi}^{\pi} g(x)\, dx \\ a_n &:= \frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\cos(nx)\, dx \\ b_n &:= \frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\sin(nx)\, dx. \end{aligned} \tag{95.1}$$

Then, we have (Parseval's identity)

$$\frac{1}{\pi}\int_{-\pi}^{\pi} g^2(x)\, dx = 2a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2). \tag{95.2}$$

Proof of Lemma: We use the orthogonality of $\cos(mx)$ and $\sin(nx)$ on $(-\pi, \pi)$:

$$\int_{-\pi}^{\pi} \cos(mx)\cos(nx)\, dx = 0 = \int_{-\pi}^{\pi} \sin(mx)\sin(nx)\, dx, \qquad \text{as long as } m \neq n.$$

$$\int_{-\pi}^{\pi} \cos(mx)\sin(nx)\, dx = 0. \tag{95.3}$$

Note also that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2(nx)\, dx = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{1+\cos(2nx)}{2}\, dx = 1 \tag{95.4a}$$

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2(nx)\, dx = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{1-\sin(2nx)}{2}\, dx = 1 \tag{95.4b}$$

Some remarks on Fourier series and Fourier coefficients:

Suppose

$$g(x) = a_0 + \sum_{n=1}^{\infty}(a_n\cos(nx) + b_n\sin(nx)) \tag{$*$}$$

is assumed to be convergent for all $x \in (-\pi, \pi)$. What are a_n, b_n? The a_n, b_n will be given by (95.1) owing to the orthonormality of the $\cos(mx)$ and the $\sin(nx)$.

Conversely, given an integrable g, define a_n, b_n via (95.1) and consider the series $(*)$. So, we write

$$g \sim a_0 + \sum_{n=1}^{\infty}(a_n\cos(nx) + b_n\sin(mx))$$

because it is not clear whether (i) the above series converges, and if so (ii) whether it conveges to $g(x)$.

It is known that:

(I) If g is continuous, then the Fourier series is $(C,1)$-summable g, i.e., if s_n denotes the n^{th} partial sum of $(*)$, then

$$\frac{s_1 + s_2 + \cdots + s_n}{n} \to g.$$

(II) $a_n, b_n \xrightarrow{n\to\infty} 0$. (Riemann-Lebesgue)

(III) If $g \in C^1[-\pi,\pi]$, then the Fourier series converges to g on $(-\pi,\pi)$.

Thus, for $g(x) \in C^1[-\pi,\pi]$, we have that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} g^2(x)\,dx = \frac{1}{\pi}\int_{-\pi}^{\pi}\left(a_0 + \sum_{n=1}^{\infty} a_n\cos(nx) + b_n\sin(nx)\right)^2 dx.$$

Because of the orthonormality, the cross-terms disappear. What is left is:

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi} g^2(x)\,dx &= \frac{1}{\pi}\int_{-\pi}^{\pi} a_0^2\,dx + \sum_{n=1}^{\infty}\frac{a_n^2}{\pi}\int_{-\pi}^{\pi}\cos^2(nx)\,dx + \sum_{n=1}^{\infty}\frac{b_n^2}{\pi}\int_{-\pi}^{\pi}\sin^2(nx)\,dx \\ &= 2a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2). \ \square\end{aligned}$$

Proof of Theorem 95: We will apply the lemma with

$$g(x) = x^k \tag{95.6}$$

for $k = 1, 2, \ldots$ and establish the theorem by induction on k.

Case 1: $(k = 1)$

$$a_0 = a_n = 0 \text{ for } n = 1, 2, \ldots. \tag{95.7}$$

because $x\cos(nx)$ is an odd function.

$$\begin{aligned}b_n = b_n(1) &= \frac{1}{\pi}\int_{-\pi}^{\pi} x\sin(nx)\,dx \\ &= \frac{-x\cos(nx)}{\pi n}\Big|_{-\pi}^{\pi} + \frac{1}{\pi n}\int_{-\pi}^{\pi}\cos(nx)\,dx \\ &= \frac{(-1)^{n+1}\cdot 2}{n}.\end{aligned} \tag{95.8}$$

So, by Parserval (Lemma 95-L), we have that

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi} x^2\,dx &= \frac{2\pi^2}{3} \\ &= \sum_{n=1}^{\infty} b_n^2 \\ &= \sum_{n=1}^{\infty}\frac{4}{n^2} \\ &= 4\zeta(2).\end{aligned} \tag{95.9}$$

Therefore,

$$\zeta(2) = \frac{\pi^2}{6}. \tag{95.10}$$

Case 2: $(k = 2)$

Here, we have

$$b_n = 0, \text{ for } n = 1, 2, \ldots. \tag{95.11}$$

Now,

$$\begin{aligned} a_0(2) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} x^2\, dx \\ &= \int_{-\pi}^{\pi} \frac{\pi^2}{3}. \end{aligned} \tag{95.12a}$$

Also,

$$\begin{aligned} a_n(2) &= \frac{1}{\pi}\int_{-\pi}^{\pi} x^2 \cos(nx)\, dx \\ &= \left|\frac{x^2\sin(nx)}{\pi n}\right|_{-\pi}^{\pi} - \frac{2}{\pi}\int_{-\pi}^{\pi}\frac{x\sin(nx)}{n}\, dx \\ &= \left|\frac{2\cos(nx)}{\pi n^2}\right|_{-\pi}^{\pi} - \frac{2}{\pi n^2}\int_{-\pi}^{\pi}\cos(nx)\, dx \\ &= \frac{(-1)^n}{4}n^2. \end{aligned} \tag{95.12b}$$

Therefore,

$$\begin{aligned} \frac{1}{\pi}\int_{-\pi}^{\pi} x^4\, dx &= \frac{2\pi^4}{5} \\ &= 2a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2) \\ &= \frac{2\pi^4}{9} + 16\sum_{n=1}^{\infty}\frac{1}{n^4}. \end{aligned} \tag{95.13}$$

Hence,

$$\zeta(4) = \frac{1}{6}\left(\frac{2\pi^4}{5} - \frac{2\pi^4}{9}\right) = \frac{\pi^4}{90}. \tag{95.14}$$

Case 3: $(k = 3)$

First note that

$$a_n(3) = 0, \text{ for } n = 0, 1, \ldots. \tag{95.15}$$

Also

$$\begin{aligned} b_n(3) &= \frac{1}{\pi}\int_{-\pi}^{\pi} x^3\sin(nx)\, dx \\ &= \left|\frac{-x^3\cos(nx)}{\pi n}\right|_{-\pi}^{\pi} + \frac{3}{\pi n}\int_{-\pi}^{\pi} x^2\cos(nx)\, dx \\ &= \frac{2\pi^2(-1)^{n+1}}{n} + \frac{3}{n}a_n(2) \\ &= \frac{2\pi^2(-1)^{n+1}}{n} + \frac{12(-1)^n}{n^3}. \end{aligned} \tag{95.16}$$

Hence,

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi} x^6\,dx &= \frac{2\pi^6}{7}\\ &= \sum_{n=1}^{\infty} b_n^2\\ &= \sum_{n=1}^{\infty}\left(\frac{2\pi^2}{n}-\frac{12}{n^3}\right)^2\\ &= 4\pi^4\underbrace{\sum_{n=1}^{\infty}\frac{1}{n^2}}_{\zeta(2)} - 48\pi^2\underbrace{\sum_{n=1}^{\infty}\frac{1}{n^4}}_{\zeta(4)} + 144\underbrace{\sum_{n=1}^{\infty}\frac{1}{n^6}}_{\zeta(6)}.\end{aligned}$$

So,

$$144\zeta(6) = \pi^6\left(\frac{2}{7}-\frac{2}{3}+\frac{8}{15}\right) = \frac{16\pi^6}{105}.$$

Thus,

$$\zeta(6) = \frac{\pi^6}{945}. \tag{95.17}$$

More generally, by induction on k, we can show:

For k odd:

$$b_n(k) = \frac{1}{n} f_k\left(\pi^2, \frac{1}{n^2}\right)$$

where $f_k(u,v)$ is a homogeneous polynomial in u, v of degree $(k-1)/2$.

For k even:

$$\begin{aligned}a_0(k) &= \frac{\pi^k}{k+1}\\ a_n(k) &= g_k\left(\pi^2, \frac{1}{n^2}\right),\end{aligned}$$

where $g_k(u,v)$ is a homogeneous polynomial in u, v of degree $k/2$.

Hence when k is odd:

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi} x^k dx &= \frac{\pi^{2k}}{2k+1}\\ &= \sum_{n=1}^{\infty} b_n^2(k)\\ &= \sum_{n=1}^{\infty}\frac{1}{n^2} f_k^2\left(\pi^2, \frac{1}{n^2}\right)\\ &= \sum_{j=0}^{k} c_j \pi^{2j}\zeta(2k-2j).\end{aligned}$$

And so in the odd case

$$\zeta(2k) = \frac{p_k}{q_k}\pi^{2k}.$$

The above part for the k is even case is analogous. □

Chapter 12

The Transcendence Theorems of Lindemann and Weierstrass

The method of to prove the transcendence of e and the proof of the transcendence of π due to Lindemann, were used by Lindemann to prove the following more general result.

Theorem 97: (Lindemann, 1882) If $\alpha \neq 0$ is algebraic, then e^{α} is transcendental.

Corollary 97.1: π is transcendental.

> **Proof:** If π were algebraic, then $i\pi$ would also be algebraic. Then, $e^{i\pi} = -1$ would be transcendental, which is a contradiction. Hence π is transcendental. Consequently, the problem of squaring the circle is settled in the negative.

Recall the proof of the transcendence of π:

> We assumed that π was algebraic and hence $i\pi$ was algebraic, and we can let $\alpha_2, \ldots, \alpha_n$ be the set of all (other) conjugates of $i\pi$. Then note that
>
> $$\left(1 + e^{i\pi}\right)\left(1 + e^{\alpha_2}\right) \cdots \left(1 + e^{\alpha_n}\right) = 0.$$
>
> Using the method of Hermite, we derived a contradiction.

We may view $i\pi$ as $\log(-1)$. From this view, we have the following corollary.

Corollary 97.2: If $\alpha \neq 0, 1$ is algebraic, then $\log(\alpha)$ is transcendental for any branch of $\log(\alpha)$.

Corollary 97.3: If $\alpha \neq 0$ is algebraic, then $\cos(\alpha)$, $\sin(\alpha)$, $\tan(\alpha)$, $\csc(\alpha)$, $\sec(\alpha)$, $\cot(\alpha)$ are all transcendental as well.

> **Proof:** Write
>
> $$\cos(\alpha) = \frac{e^{i\alpha} + e^{-i\alpha}}{2} =: \beta/2$$
>
> and assume toward a contradiction that β is algebraic. Then,
>
> $$2\cos(\alpha) = e^{i\alpha} + \frac{1}{e^{i\alpha}} = \beta$$
>
> and hence
>
> $$\left(e^{i\alpha}\right)^2 - \beta e^{i\alpha} + 1 = 0. \tag{97.1}$$

Therefore, $e^{i\alpha}$ is the root of a quadratic equation with algebraic coefficients. Hence, $e^{i\alpha}$ is algebraic, because from (97.1) we may a construct a polynomial with integer coefficients of (possibly much) larger degree for which $e^{i\alpha}$ is a root. To construct this polynomial, consider

$$(e^{i\alpha})^2 - \beta' e^{i\alpha} + 1 \tag{97.2}$$

as β' runs through all conjugates of β, and multiply all such expressions in (97.2) by the expression in (97.1) to get this polynomial in $\mathbb{Q}[x]$ for which $e^{i\alpha}$. This is a contradiction, and so $\cos(\alpha)$ is transcendental.

The transcendence of $\sin(\alpha)$ follows from the equation $\cos^2(\alpha) + \sin^2(\alpha) = 1$. Similarly for $\tan(\alpha)$ using the formula $\tan^2(\alpha) = \sec^2(\alpha) - 1$. □

Theorem 98 (Lindemann's Theorem): If $\alpha_1, \ldots, \alpha_n$ are algebraic numbers which are linearly independent over $\mathbb{Q}$, then $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are algebraically independent.

Corollary: Theorem 97. **Proof:** Apply $n = 1$.

Lindemann sketched a proof. In establishing **Theorem 98** rigorously, Weierstrass strengthened it as follows:

Theorem 99: (Lindemann-Weierstrass) If $\alpha_1, \ldots, \alpha_n$ are distinct algebraic numbers, then for non-zero algebraic numbers $\beta_1, \ldots, \beta_n$, we have that

$$\beta_1 e^{\alpha_1} + \cdots + \beta_n e^{\alpha_n} \neq 0. \tag{99.1}$$

Remark 1: (99.1) is equivalent to saying that $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are linearly independent over the field of algebraic numbers.

Remark 2: If all of the α_i are nonzero, then we apply the theorem with $\alpha_{n+1} = 0$, and we now have $n + 1$ distinct algebraic numbers. In this case (99.1) is strengthened to the following:

$$\beta_1 e^{\alpha_1} + \cdots + \beta_n e^{\alpha_n} \neq \beta_{n+1} \tag{f}$$

or any algebraic β_{n+1}.

At first glance, it appears that **Theorem 98** is stronger because it implies algebraic independence, whereas **Theorem 99** only implies linear independence over the set of algebraic elements. In fact, **Theorem 99** implies **Theorem 98**, and we will now prove this:

Proof: If $\gamma_1, \ldots, \gamma_n$ are algebraically dependent, then there exists a polynomial

$$p(x_1, \ldots, x_n) \in \mathbb{A}[x_1, \ldots, x_n]$$

such that

$$p(\gamma_1, \ldots, \gamma_n) = 0.$$

Given $p \in \mathbb{A}[x_1, \ldots, x_n]$, write

$$p(x_1, \ldots, x_n) = \sum_{(j_1, \ldots, j_n) \in \mathbb{Z}_{\geq 0}} \beta_{j_1, \ldots, j_n} x_1^{j_1} \cdots x_n^{j_n}.$$

If, according to the hypothesis of **Theorem 98**, we have that $\alpha_1, \ldots, \alpha_n$ are linearly independent over $\mathbb{Q}$ (algebraic numbers) and hence linearly independent over $\mathbb{Z}$, then

$$e^{j_1\alpha_1 + \cdots + j_n\alpha_n} = x^{j_1} \cdots x^{j_n} \Big| \begin{Bmatrix} x_1 = e^{\alpha_1} \\ \vdots \\ x_n = e^{\alpha_n} \end{Bmatrix} \tag{99.5}$$

Since $\alpha_1, \ldots, \alpha_n$ are linearly independent over $\mathbb{Z}$, the values $j_1\alpha_1 + \cdots + j_n\alpha_n$ would all be different for distinct tuples $(j_1, \ldots, j_n)$.

So, by **Theorem 99**, (99.5), and (99.4),

$$p(e^{\alpha_1}, \ldots, e^{\alpha_n}) \neq 0.$$

Hence, $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are algebraically independent. □

Now we actually prove **Theorem 99**.

Theorem 99: (Lindemann-Weierstrass) If $\alpha_1, \ldots, \alpha_n$ are distinct algebraic numbers, then for non-zero algebraic numbers $\beta_1, \ldots, \beta_n$, we have that

$$\beta_1 e^{\alpha_1} + \cdots + \beta_n e^{\alpha_n} \neq 0. \tag{99.1}$$

Proof: Suppose that the theorem is false. Then, there exist distinct algebraic $\alpha_1, \ldots, \alpha_n$, and algebraic $\beta_1, \ldots, \beta_n$ such that

$$\beta_1 e^{\alpha_1} + \cdots + \beta_n e^{\alpha_n} = 0. \tag{99.2}$$

We may assume without loss of generality that in (99.2) all β_i are nonzero because otherwise, we only consider the nonzero β_i.

Next we form all possible expressions

$$\beta_1' e^{\alpha_1} + \cdots + \beta_n' e^{\alpha_n} \tag{99.3}$$

where $\beta_1', \ldots, \beta_n'$ run independently through all conjugates of $\beta_1, \ldots, \beta_n$, and we multiply all such expressions and this will be 0. We end up with terms of the form

$$e^{(\sum j_I \alpha_i)} \tag{99.4}$$

where the $j_i \in \mathbb{N}$ and their coefficients are rational. We can then multiply this by the least common multiple and obtain an expression with integer coefficients.

Some of the values $\sum j_i \alpha_i$ could be the same. So we group the exponents with same values $\sum j_i \alpha_i$ and calculate each such coefficients. These will be our new αs. So we can assume that in (99.2) we are dealing with coefficients β_i which are (rational) integers. (Recall that the rational integers are just the usual integers. We use the word rational to distinguish from the algebraic integers, which we will mention later.) We will get a contradiction in this case. This is sufficient.

Next, since each α_i is algebraic, α_i is the root of a minimal polynomial (with integer coefficients). Thus, $\alpha_1, \ldots, \alpha_n$ are collectively the roots of a polynomial (which may not be irreducible). Let $\alpha_{n+1}, \ldots, \alpha_N$ be the remaining roots of this (monstrous) polynomial. Thus, for any α_i, the list $\alpha_1, \ldots, \alpha_N$ will contain all of the conjugates of α_2.

Now consider the expression

$$\prod (\beta_1 e^{\alpha_{k_1}} + \cdots + \beta_N e^{\alpha_{k_N}}) = 0 \tag{99.5}$$

where $(k_1, k_2, \ldots, k_N)$ run through all permutations of $1, 2, \ldots, N$, and with

$$\beta_{n+1} = \beta_{n+2} = \cdots = \beta_N = 0. \tag{99.6}$$

The product in (99.5) may be expanded and expression as a sum of terms of the form

$$e^{h_1\alpha_1 + \cdots + h_N \alpha_N} \tag{99.7}$$

where the $h_2 \geq 0$ are integers satisfying

$$h_1 + h_2 + \cdots + h_N = N!. \tag{99.8}$$

The expressions

$$h_1\alpha_1 + \cdots + h_N\alpha_N \tag{99.9}$$

will run through a full set of conjugates. These will be our new αs, with the property that he coefficient of e^{α} will be the same as that of $e^{\alpha'}$ where α, α' are expressions as in (99.9) and α, α' are conjugates. This is now our new expression in (99.2). So, after rearranging the terms, this can be put in the following form: there exist integers

$$0 = n_0 < n_1 < n_2 < \cdots < n_r = n \tag{99.10}$$

such that

$$\alpha_{n_t+1}, \alpha_{n_t+2}, \ldots, \alpha_{n_{t+1}} \tag{99.11}$$

is a full set of conjuates and their coefficients are all equal, i.e.,

$$\beta_{n_t+1} = \beta_{n_t+2} = \cdots = \beta_{n_{t+1}}. \tag{99.12}$$

Remark: At least one of the βs will be non-zero. This is realized by imposing an order on the complex numbers $\mathbb{C}$ by declaring

$$z_1 < z_2, \text{ if } \begin{cases} \mathrm{Re}(z_1) < \mathrm{Re}(z_2), \text{ or} \\ \mathrm{Re}(z_1) = \mathrm{Re}(z_2) \text{ and } \mathrm{Im}(z_1) = \mathrm{Im}(z_2) \end{cases}. \tag{99.13}$$

Since $\alpha_1, \ldots, \alpha_n$ are algebraic numbers, there exists an integer ℓ such that $\ell\alpha_1, \ldots, \ell\alpha_n$ are algebraic integers.

Algebraic Numbers & Algebraic Integers: Let α algebraic be a root of

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n \tag{99.14}$$

which is irreducible with $\gcd(a_0, \ldots, a_n) = 1$. Therefore $p(\alpha) = 0$. We say that α is an algebraic integers if $a_n = 1$. Let $\beta = a_n\alpha$. Therefore,

$$p\left(\frac{\beta}{a_n}\right) = a_0 + a_1\frac{\beta}{a_n} + a_2\frac{\beta^2}{a_2^2} + \cdots + a_n\frac{\beta^n}{a_n^n} = 0. \tag{99.15}$$

Therefore, multiplying through by a_n^{n-1}:

$$a_0 a_n^{n-1} + \cdots + a_{n-2}a_{n-1}\beta^{n-2} + a_{n-1}\beta^{n-1} + \beta^n = 0. \tag{99.16}$$

Therefore, β is the root of a monic polynomial with integer coefficients, and hence $\beta = a_n\alpha$ is an algebraic integer.

Now consider for a large prime p (to be specified later)

$$f_i(x) = \frac{\ell^{np}\{(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n)\}^p}{x-\alpha_i} \tag{99.17}$$

for $i = 1, 2, \ldots, n$. ℓ was chosen so that $\ell\alpha_i$ is an algebraic integer for all i. Also consider the integrals

$$I_i(z) = \int_0^z e^{z-w} f_i(w)\, dw \tag{99.18}$$

as in Hermite's proof of the transcendence of e. Also consider the expressions

$$J_i = \beta_1 I_i(\alpha_1) + \beta_2 I_i(\alpha_2) + \cdots + \beta_n I_i(\alpha_n) \tag{99.19}$$

for $i = 1, 2, \ldots, n$.

By repeated integration by parts, we have

$$J_i = \left(\sum_{i=1}^n \beta_i e^{\alpha_i}\right)\left(\sum_{j=0}^m f_i^{(j)}(0)\right) - \sum_{j=0}^m\sum_{k=1}^n \beta_k f_i^{(j)}(\alpha_k) = -\sum_{j=0}^m\sum_{k=1}^n \beta_k f_i^{(j)}(\alpha_k) \tag{99.20}$$

in our new (99.2), where $m = np - 1 = \deg(f_i)$.

The polynomials $f_i(x)$ in (99.17) have algebraic coefficients. We do not have a full set of conjugates because we are dividing by the factor $x - \alpha_i$. Hence note that each value $f_i^{(j)}(\alpha_k)$ is an algebraic integer with the property:

$$p! \mid f_i^{(j)}(\alpha_j), \text{ for all } j, i \text{ except for } j = p-1,\ k = i \tag{99.21a}$$

and in fact

$$f_i^{(p-1)}(\alpha_i) = \ell^{np}(p-1)! \prod_{\substack{k=1 \\ k\neq i}}^{n} (\alpha_i - \alpha_k)^p \,. \tag{99.21b}$$

Note that when we say that $p!$ divides as above, we mean as an algebraic integer. That is, $f_i^{(j)}(\alpha_k)$ is $p!$ times an algebraic integer.

So, if p is large enough, then

$$(p-1)! \mid f_i^{(p-1)}(\alpha_i), \quad \text{but} \quad p! \nmid f_i^{(p-1)}(\alpha_i). \tag{99.22}$$

Thus, J_i is a multiple of $(p-1)!$ but not of $p!$, i.e., J_i is a nonzero multiple of $(p-1)!$.

Next, rewrite the expression on the right in (99.20) as

$$J_i = -\sum_{j=0}^{m} \sum_{t=0}^{r-1} \beta_{n_{t+1}} \left(f_i^{(j)}(\alpha_{n_t+1}) + \cdots + f_i^{(j)}(\alpha_{n_{t+1}}) \right) \tag{99.23}$$

for all $i = 1, 2, \ldots, n$. From (99.23) it follows that

$$J_1 J_2 \cdots J_n \in \mathbb{Z} \smallsetminus \{0\}. \tag{99.24a}$$

Therefore,

$$|J_1 J_2 \cdots J_n| \geq ((p-1)!)^n \,. \tag{99.24b}$$

On the other hand, we have trivially from (99.18) and (99.19) that

$$|J_i| \leq \sum_{k=1}^{n} |\beta_k| e^{|\alpha_k|} \widehat{f}(|\alpha_k|)|\alpha_k| \leq C^p \tag{99.25}$$

for some large constant C which depends on ℓ and the $|\alpha_i$ but not on p. This is a crude upper bound. Recall that $\widehat{f}$ as in Hermite's proof is the polynomial f with the absolute value function applied to each coefficient of f.

Hence,

$$|J_1 J_2 \cdots J_n| \leq C^{np}. \tag{99.26}$$

This is incompatible with the lower bound in (99.24). Thus we have a contradiction. □

Chapter 13

The Gelfond-Schneider Theorem

Hilbert's 7^{th} Problem:

(1) In an isosceles triangle, if the ratio of the base angle to the vertex angle is algebraic irrational, then the ratio between the base length and the side length is transcendental.

(2) Is a^b transcendental when $a \notin \{0, 1\}$ is algebraic and b is algebraic irrational? ("Hilbert's Conjecture" is that the answer is the affirmative. This was confirmed by the **Gelfond-Schneider Theorem**.) We will call this **Conjecture 100**.

Reformulation of Conjecture 100:

Write $\alpha^\beta = \gamma$ as $\beta \log(\alpha) = \log(\gamma)$, and therefore,

$$\frac{\log(\gamma)}{\log(\alpha)} = \log_\alpha(\gamma) = \beta.$$

Equivalently, if $\alpha \neq 0, 1$ is algebraic and γ is algebraic, then

$$\frac{\log(\gamma)}{\log(\alpha)} = \log_\alpha(\gamma)$$

is either rational or transcendental.

Theorem 100′: (Gelfond, 1929) If α, γ are algebraic with $\alpha \neq 0, 1$, then $\dfrac{\log(\gamma)}{\log(\alpha)}$ cannot be an imaginary quadratic irrational.

Consequence: e^π is transcendental. This is true because $e^\pi = \left(e^{i\pi}\right)^{-i} = (-1)^{-i}$.

Theorem 100″: (Kuzmin, 1930) If α, γ are algebraic, with $\alpha \neq 0, 1$, then $\dfrac{\log(\gamma)}{\log(\alpha)}$ cannot be a real quadratic irrational. **Consequence:** $2^{\sqrt{2}}$ (which is known as the Hilbert Number) is transcendental.

Theorem 100: (Gelfond-Schneider, 1934) **Conjecture 100** is true.

Observation: Conjecture 100(2) implies Conjecture 100(1).

Proof: To see this, consider an isosceles triangle with base angle θ and top angle φ. Recall that $2\theta + \varphi = \pi$, which is transcendental. Assume that

$$\varphi = \alpha\theta, \tag{100.2}$$

for α an algebraic irrational. Now,

$$\begin{aligned} 2\theta + \alpha\theta &= \pi \\ \theta(2+\alpha) &= \pi \\ \theta &= \frac{\pi}{2+\alpha} = \beta\pi, \end{aligned}$$

for some β which is an algebraic irrational.

Recall that

$$\begin{aligned} \frac{\text{base}}{\text{side}} &= 2\cos(\theta) \\ &= e^{i\theta} + e^{-i\theta} \\ &= e^{i\beta\pi} + \frac{1}{e^{i\beta\pi}} \\ &= (-1)^{\beta} + (-1)^{-\beta}. \end{aligned}$$

By the **Gelfond-Schneider Theorem**, we have that $(-1)^{\beta}$ is transcendental. Hence,

$$(-1)^{\beta} + \frac{1}{(-1)^{\beta}}$$

is also transcendental.

Remark: We reformulated the **Gelfond-Schneider Theorem** as follows: If α, β are algebraic and not 0 or 1, and if their ratio is not rational, then their ratio is transcendental.

Observe that if $\mathbb{F}$ is a subfield of $\mathbb{C}$, then $x_1, x_2 \in \mathbb{C}$ are linearly independent over $\mathbb{F}$ if and only if neither x_1 nor x_2 is a multiple of the other (with the multiple coming from $\mathbb{F}$). So, this yields another version of the **Gelfond-Schneider Theorem**.

Theorem 100-R: If α_1, α_2 are algebraic numbers such that $\log(\alpha_1)$ and $\log(\alpha_2)$ are linearly independent over $\mathbb{Q}$, then $\log(\alpha_1)$ and $\log(\alpha_2)$ are linearly independent over the field of algebraic numbers, i.e.

$$\beta_1 \log(\alpha_1) + \beta_2 \log(\alpha_2) \neq 0 \tag{100.5}$$

for all β_1, β_2 which are algebraic numbers not both zero.

Remark: Baker's theorems deal with an n-variable extension of (100.5), published in 1966 and 1967 in Mathematika.

Theorem 101: (Baker) Let $\alpha_1, \ldots, \alpha_n$ be non-zero algebraic numbers such that $\log(\alpha_1), \ldots, \log(\alpha_n)$ are linearly independent over $\mathbb{Q}$. Then, $1, \log(\alpha_1), \ldots, \log(\alpha_n)$ are linearly independent over the field of algebraic numbers.

Remark: In the case $n = 2$, this theorem yields

$$\beta_1 \log(\alpha_1) + \beta_2 \log(\alpha_2) \neq \gamma, \tag{101.1}$$

where γ is algebraic, i.e.

$$\alpha_1^{\beta_1} \neq \alpha_2 e^{\gamma}, \tag{101.2}$$

for γ algebraic. This strengthening of the **Gelfond-Schneider Theorem** will play a role subsequently in our discussion of transcendental functions.

Theorem 101′: (Baker) If $\alpha_1, \ldots, \alpha_n$ are algebraic numbers, and if $\beta_1, \ldots, \beta_n$ are algebraic such that

$$\beta_1 \log(\alpha_1) + \cdots + \beta_n \log(\alpha_n) \neq 0, \tag{101.3}$$

then the expression actually is transcendental.

Remark: This theorem is equivalent to saying that if $\beta_0 \neq 0$ is algebraic, then

$$\beta_0 + \beta_1 \log(\alpha_1) + \cdots + \beta_n \log(\alpha_n) \neq 0. \tag{101.4}$$

We can deduce (101.4) from **Theorem 101**, and so **Theorem 101** $\Longrightarrow$ **Theorem 101′**. We now prove this.

Proof that 101 $\Rightarrow$ 101′: Clearly, (101.4) holds for $n = 0$. So, assume that (101.4) holds for all $n < m$ where $m \in \mathbb{Z}^+$. Consider the expression in (101.4) for $n = m$. If $\log(\alpha_1), \ldots, \log(\alpha_n)$ are linearly independent over $\mathbb{Q}$, then (101.4) follows from **Theorem 101**. Hence, we only need to consider the case where $\log(\alpha_1), \ldots, \log(\alpha_n)$ are linearly dependent over $\mathbb{Q}$. Thus let

$$\rho_1 \log(\alpha_1) + \cdots + \rho_m \log(\alpha_m) = 0 \tag{101.5}$$

where $\rho_i \in \mathbb{Q}$ and ρ_i not all 0. Suppose

$$\rho_r \neq 0. \tag{101.6}$$

Then,

$$\begin{aligned}
\rho_r \left(\beta_0 + \beta_1 \log(\alpha_1) + \cdots + \beta_m \log(\alpha_m)\right) &= \rho_r \left(\beta_0 + \beta_1 \log(\alpha_1) + \cdots + \beta_m \log(\alpha_m)\right) - 0 \\
&= \rho_r \left(\beta_0 + \beta_1 \log(\alpha_1) + \cdots + \beta_m \log(\alpha_m)\right) - \beta_r \cdot 0 \\
&= \rho_r \left(\beta_0 + \beta_1 \log(\alpha_1) + \cdots + \beta_m \log(\alpha_m)\right) \\
&\qquad - \beta_r \left(\rho_1 \log(\alpha_1) + \cdots + \rho_m \log(\alpha_m)\right) \\
&= \beta_0' + \beta_1' \log(\alpha_1) + \cdots + \beta_m' \log(\alpha_m)
\end{aligned} \tag{101.7}$$

where

$$\beta_0' = \rho_r \beta_0 \neq 0, \quad \text{and} \quad \beta_j' = \rho_r \beta_j - \beta_r \rho_j \tag{101.8}$$

with $\beta_r' = 0$.

Thus, the expression in (101.7) is of the type as in (101.4), but with $n < m$. Thus by the induction hypothesis, the expression on the right in (101.7) is not equal to 0. So the expression on the left of (101.7) is not equal to zero. □

Theorem 101″: If $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ are non-zero algebraic numbers, then $e^{\beta_0} \alpha_1^{\beta_1} \cdots \alpha_n^{\beta_n}$ is transcendental.

Proof: If $e^{\beta_0} \alpha_1^{\beta_1} \cdots \alpha_n^{\beta_n} = \alpha_{n+1}$ is algebraic and nonzero, then

$$\beta_1 \log(\alpha_1) + \cdots + \beta_n \log(\alpha_n) - \log(\alpha_{n+1}) = \beta_0 \neq 0. \tag{101.9}$$

This fails by **Theorem 101′** for $n + 1$. □

Theorem 101‴: Let $\alpha_1, \ldots, \alpha_n$ be algebraic and not equal to 0 or 1, and let $\beta_1, \ldots, \beta_n$ be algebraic such that $1, \beta_1, \ldots, \beta_n$ are linearly independent over $\mathbb{Q}$. Then, $\alpha_1^{\beta_1} \cdots \alpha_n^{\beta_n}$ is transcendental.

Remark: When $n = 1$, the hypothesis that $1, \beta_1$ are linearly independent over $\mathbb{Q}$ is the same as saying β_1 is irrational. This special case is the **Gelfond-Schneider Theorem**.

Proof: We will show that for any choice of algebraic numbers $\alpha_1, \ldots, \alpha_n$ not equal to 0 or 1, and any choice $\beta_1, \ldots, \beta_n$ of algebraic numbers, that

$$\beta_1 \log(\alpha_1) + \cdots + \beta_n \log(\alpha_n) \neq 0. \tag{101.10}$$

This needs to be shown for all n, since if (100.10) is true, then **Theorem** $'''$ follows, because if

$$\alpha_1^{\beta_1} \cdots \alpha_n^{\beta_n} = \alpha_{n+1} \tag{101.11}$$

is algebraic, then $\beta_1 \log(\alpha_1) + \cdots + \beta_n \log(\alpha_n) - \log(\alpha_{n+1} = 0$, which violates (101.10) for $n+1$.

Note that $\beta_1, \ldots, \beta_n$ linearly independent over $\mathbb{Q}$ with $\beta_{n+1} = -1$ is the same as $1, \beta_1, \ldots, \beta_n$ linearly independent over $\mathbb{Q}$. We will establish (101.10) by induction on n. It is clearly true for $n = 1$, so let (101.10) be true for $n < m$. Consider now $n = m$. If $\log(\alpha_1), \ldots, \log(\alpha_n)$ are linearly independent over $\mathbb{Q}$, then (101.10) holds because of **Theorem 101**. So, we only consider the case where $\log(\alpha_1), \ldots, \log(\alpha_n)$ are linearly dependent over $\mathbb{Q}$, i.e., there exists $\rho_1, \ldots, \rho_m \in \mathbb{Q}$ such that

$$\rho_1 \log(\alpha_1) + \cdots + \rho_m \log(\alpha_m) = 0 \tag{101.12}$$

with $\rho_r \neq 0$. Now consider that

$$\rho_r \left(\beta_1 \log(\alpha_1) + \cdots + \beta_m \log(\alpha_m)\right) = \beta_1' \log(\alpha_1) + \cdots + \beta_m' \log(\alpha_m) \tag{101.13}$$

with $\beta_r = 0$. The result follows by induction. $\square$

Two Striking Consequences:

Corollary 101-1: If α is algebraic and $\beta \neq 0$ is algebraic, then $e^{\pi\alpha+\beta}$ is transcendental.

Proof: Set $e^{\pi\alpha+\beta} = \gamma$, so that

$$\pi\alpha + \beta = \log(\gamma). \tag{101.14}$$

If γ is algebraic, rewrite (101.14) as

$$i\pi\alpha - i\log(\gamma) = -i\beta,$$

i.e.,

$$\alpha \log(-1) - i\log(\gamma) = -i\beta. \tag{101.15}$$

This violated **Theorem 101′**. Hence γ is transcendental.

Corollary 101-2: If $\alpha \neq 0$ is algebraic, then $\pi + \log(\alpha)$ is transcendental.

Effective Estimates for Linear Forms of Algebraic Numbers (due to Baker)

Definition: Given an algebraic number α, we can find a polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ of least degree such that $\gcd\{a_0, \ldots, a_n\} = 1$ and $p(\alpha) = 0$. The height of α is

$$h(\alpha) := \max_{0 \leq i \leq n} |a_i|.$$

Given

$$\Lambda = \beta_1 \log(\alpha_1) + \beta_2 \log(\alpha_2) + \cdots + \beta_n \log(\alpha_n)$$

with $\Lambda \neq 0$, Baker obtained effective lower bounds for $|\Lambda - \beta|$, i.e., how close to an algebraic number Λ can get. He found

$$|\Lambda - \beta| \geq C(\Lambda, \beta) > 0,$$

where $C(\lambda, \beta)$ was an effectively computable constant in terms of (i) the degree of α_i, β_i and (ii) the heights of α_i, β_i. Such bounds showed that certain classes of Diophantine Equations had only a finite number of solutions by obtaining a bound on the size of any solutions. These bounds were often triple order exponents, e.g. $10^{10^{500}}$.

Catalan's Conjecture: (Mihăilescu's Theorem)

Catalan conjectured that 8 and 9 are the only two consecutive powers. This was communicated in 1844 in a letter to Crelle's Journal.

Theorem 102: (Robert Tidjeman, 1976) The Diophantine Equation

$$x^m - y^n = 1 \tag{102.1}$$

for $m, n, x, y > 1$ integers has only a finite number of solutions. A large solution to (102.1) implies that

$$m\log(x) - n\log(y)$$

is very small. In 2002, Preda Mihăilescu (who was at the time not employed) solved Catalan's conjecture. We will present a special case.

Theorem 102-S: For $x, y > 1$, $[x = 2,\ y = 3]$ is the only solution to the Diophantine equation

$$3^x - 2^y = 1,. \tag{102.2}$$

Remark: This result was known as early as 1343 (Gersonides).

Proof: Note that (102.2) implies

$$2^y \equiv -1 \pmod{3^x}$$

and so

$$2^{2y} \equiv 1 \pmod{3^x}. \tag{102.3}$$

Clearly, 2 is a primative root of 3, i.e. $(\mathbb{Z}/3\mathbb{Z})^\times = \{1, 2\} = \langle 2\rangle$. It turns out that 2 is also a primative root of 3^2, since in $(\mathbb{Z}/9\mathbb{Z})^\times$, we have

$$\langle 2\rangle = \{2, 4, 8, 7, 5, 1\} = (\mathbb{Z}/9\mathbb{Z})^\times.$$

In fact, since 2 is a primative root of 3^2, it is a primative root of 3^x for all $x \in \mathbb{Z}^+$.

Therefore, in (102.3), we have

$$\varphi\left(3^x\right) = 3^{x-1} \cdot 2 \tag{102.4}$$

must divide $2y$. Thus $3^{x-1} \mid y$ and so

$$y \geq 3^{x-1}. \tag{102.5}$$

Hence

$$2^y \geq 2^{3^{x-1}}. \tag{102.6}$$

This is far too big to fit into (102.2) unless x is very small. Indeed, $x = 2$ is the only one that works. □

Chapter 14

Transcendence Degree and Transcendental Functions

14.1 Definitions

Definition: The transcendence degree of a field extension L over K is the largest cardinality of an algebraically independent subset of L over K.

Definition: A subset S of L is a transcendence basis for L over K if the elements of S are algebraically independent over K and if L is an algebraic extension of $K(S)$. Every field extension has a transcendence basis, and any two transcendence bases have the same cardinality. That cardinality is called the transcendence degree of L over K.

Remark: Transcendence bases for field extensions are analogous (in a sense) to bases for vector spaces, with one important difference: In a vector space, each basis spans the whole vector space. In a field extension L, we may not have $L = K(S)$ for some transcendence basis S.

Definition: An field extension L over K is called purely transcendental if it has a transcendence basis S such that $L = K(S)$.

14.2 Schanuel's Conjecture and its Implications

Schanuel's Conjecture (1960s) If $z_1, z_2, \ldots, z_n$ are n complex numbers linearly independent over $\mathbb{Q}$, then

$$\text{TrDeg}\left[\mathbb{Q}\left(z_1, z_2, \ldots, z_n, e^{z_1}, e^{z_2}, \ldots, e^{z_n}\right) : \mathbb{Q}\right] \geq n.$$

Implication 1: Schanuel's Conjecture proves the **Lindemann-Weierstrass Theorem**. Take $z_1, \ldots, z_n$ to be algebraic and linearly independent over $\mathbb{Q}$. By Schanuel's Conjecture,

$$\text{TrDeg}\left[\mathbb{Q}\left(z_1, z_2, \ldots, z_n, e^{z_1}, e^{z_2}, \ldots, e^{z_n}\right) : \mathbb{Q}\right] \geq n.$$

Thus, $e^{z_1}, \ldots, e^{z_n}$ must be algebraically independent, and this is the **Lindemann-Weierstrass Theorem**.

Implication 2: Schanuel's Conjecture proves a stronger form of **Baker's Theorem**. Choose $\alpha_i := e^{z_i}$ for $i = 1, \ldots, n$ and hence $z_i = \log(\alpha_i)$. Choose z_i (i.e., α_i) such that $\log(\alpha_1), \ldots, \log(\alpha_n)$ are linearly independent over $\mathbb{Q}$. Then, by Schanuel's Conjecture, we have that

$$\text{TrDeg}\left[\mathbb{Q}\left(\log(\alpha_1), \ldots, \log(\alpha_n), \alpha_1, \ldots, \alpha_n\right) : \mathbb{Q}\right] \geq n.$$

Now, $\log(\alpha_1), \ldots, \log(\alpha_n)$ are algebraically independent, which is stronger than **Baker's Theorem**.

Implication 3: Schanuel's Conjecture implies the **Gelfond-Schneider Theorem**.

Implication 4: Schanuel's Conjecture implies that e and π are algebraically independent. Choose $z_1 = 1$, $z_2 = i\pi$, Then, z_1, z_2 are linearly independent over $\mathbb{Q}$. Now,

$$\text{TrDeg}\left[\{1, i\pi, e^1, -1\} : \mathbb{Q}\right] \geq 2.$$

Hence, e and $i\pi$ are algebraically independent, which implies that e and π are algebraically independent. If this is true, then this would imply that $e + \pi$ and $e\pi$ are transcendental, which is a famous open conjecture.

14.3 Transcendental Functions and their Exceptional Sets

Definition: Given a field $\mathbb{F}$ and $\mathbb{F}(z)$, the field of rational fractions with coefficients in $\mathbb{F}$, we define $g(z) \in \mathbb{F}(z)$ to be algebraic over $\mathbb{F}(z)$ if there exists $\rho_0, \rho_1, \ldots, \rho_n \in \mathbb{F}(z)$ such that

$$\rho_0(z) + \rho_1(z)g(z) + \rho_2(z)g^2(z) + \cdots + \rho_n(z)g^n(z) = 0,$$

with not all $\rho_i = 0$.

We may clear the denominators of each $\rho_i(z)$ and hence assume each $\rho_i(z) \in \mathbb{F}[z]$, i.e, each ρ_i is just a polynomial.

Definition: We say that $f(z) \in \mathbb{F}(z)$ is transcendental over $\mathbb{F}(z)$ if $f(z)$ is not algebraic over $\mathbb{F}(z)$.

Example: The functions e^z, 2^z, a^z $(a \neq 0, 1)$, $\log(z)$, $\cos(z)$ (in fact, all the trig functions), are all transcendental functions.

Theorem 102-T: The function $f(z) := e^z$ is a transcendental function over $\mathbb{C}[z]$.

Proof: Suppose that e^z is algebraic, say of degree d. So, there exists $p_0, p_1, \ldots, p_d \in \mathbb{C}(z)$ with $p_d \neq 0$ such that

$$p_0(z) + p_1(z)e^z + p_2(z)e^{2z} + \cdots + p_d(z)e^{dz} = 0. \tag{102.6}$$

Rewrite (102.6) as

$$p_d(z)e^{dz} = -p_{d-1}e^{(d-1)z} - \cdot - p_0(z), \tag{102.7}$$

and compare the sizes of both sides. Note that for $z = r > 0$, we have that

$$\text{lhs}(102.7) \geq e^{dr} \cdot cr^{\deg(p_d)}. \tag{102.8}$$

On the other hand

$$\text{rhs}(102.7) \leq ce^{(d-1)r^k} \tag{102.9}$$

where

$$k := \max_{0 \leq i \leq d-1} \deg(p_i).$$

The bounds in (102.8) and (102.9) are incompatible as $r \to \infty$. Thus, e^z is a transcendental function. □

Theorem: Every entire algebraic function is a polynomial. (Prove this as an exercise. Use ideas similar to **Theorem 102-T**, something about the Maximum-Modulus Theorem?)

Theorem: If $g(z)$ is algebraic over $\mathbb{Q}(z)$, then $g(\alpha)$ is algebraic for every algebraic α.

Problem: Given $f(z)$ transcendental over $\mathbb{C}(z)$, when is $f(\alpha)$ transcendental given α algebraic? I.e., determine

$$E_f := \{\alpha \in \mathbb{A} \mid f(\alpha) \text{ is transcendental}\}.$$

We call this the exceptional set of f. This questions is due to Weierstrass.

Example: Let $f(x) := e^z$. Then, $E_f = \{0\}$. (Lindemann)

Example: Let $f(x) := 2^z$. Then, $E_f = \mathbb{Q}$. (Gelfond-Schneider)

Example: Let $f(x) := \cos(z)$. Then, $E_f = \{0\}$.

Example: Let $f(x) := \log(z)$. Then, $E_f = \{0, 1\}$.

Example: Let $f(x) := e^{e^z}$. Then, $E_f = \emptyset$. Follows from Schanuel's Conjecture: Consider $z_1 = \alpha$ and $z_2 = e^\alpha$ with $\alpha \neq 0$ algebraic. Clearly z_1, z_2 are linearly independent. By Schanuel's Conjecture,

$$\text{TrDeg}\left[\{\alpha, e^\alpha, e^\alpha, e^{e^\alpha}\} : \mathbb{Q}\right] \geq 2.$$

Hence e^α and e^{e^α} are algebraically independent and thus e^{e^α} is transcendental when $\alpha \neq 0$. When $\alpha = 0$, we have $f(\alpha) = e$. Hence $E_f = \emptyset$.

Example: Let $f(z) := e^{1+z\pi}$. Then, $E_f = \emptyset$.

Proof: If $z = \alpha = 0$, then $f(0) = e$, which is transcendental. Thus we now consider $z = \alpha \neq 0$. Suppose

$$f(\alpha) = e^{1+\pi\alpha} = \beta \in \mathbb{A}.$$

Then, $1 + \pi\alpha = \log(\beta)$, and so $i + i\pi\alpha = i\log(\beta)$. Therefore, $\alpha\log(-1) - i\log(\beta) = -i$. Thus, we have a nonvanishing linear combination of $\log(-1)$ and $\log(\beta)$ (which is algebraic). This violates one of the earlier theorems due ot Baker. Hence, $f(\alpha)$ must be transcendental. □

Example: let $g(z) := e^z + e^{1+z}$. Then, $E_g = \emptyset$.

Proof: If $z = \alpha = 0$, then $g(0) = 1 + e$, which is transcendental. Similarly, $g(1) = e + e^2$, which is transcendental. Thus we now consider $z = \alpha \neq 0, 1$, i.e. $0, 1 \neq 1 + \alpha$. By the **Weierstrass Theorem**, 1, e^α and $e^{1+\alpha}$ are linearly independent over $\mathbb{A}$. In particular, $e^\alpha + e^{1+\alpha}$ is not algebraic. □

Remark: It is a surprising fact that given $B \subseteq \mathbb{A}$, there exists a transcendental function f for which $E_f = B$.

We want examples of transcendental functions (which are not growing fast or are not entire).

Theorem 102-L: Let

$$f(z) := \sum_{n=0}^{\infty} z^{2^n} \tag{102.10}$$

for $|z| < 1$. Then, f is a transcendental function.

Proof: First write that $f(z)$ satisfies the functional equation

$$f(z) = \sum_{n=0}^{\infty} z^{2^n} = z + \sum_{n=1}^{\infty} z^{2^n} = z + f(z^2). \tag{102.11}$$

This will be used to show that $f(z)$ is a transcendental function.

Suppose that $f(z)$ is algebraic. Then there exist rational functions $p_0(z), p_1(z), \ldots, p_{d-1}(z)$ such that

$$f(z)^d + p_{d-1}(z)f(z)^{d-1} + \cdots + p_1(z)f(z) + p_0(z) = 0 \tag{102.12}$$

and d is minimal and so the expression in (102.12) is uniquely determined. Thus,

$$f(z^2)^d + p_{d-1}(z^2)f(z^2)^{d-1} + \cdots + p_1(z^2)f(z^2) + p_0(z^2) = 0 \tag{102.13}$$

as well.

Using $f(z^2) = f(z) - z$, we rewrite (102.13) as

$$(f(z) - z)^d + p_{d-1}(z^2)(f(z) - z)^{d-1} + \cdots = 0. \tag{102.14}$$

In (102.14), expand each term $(f(z) - z)^j$ by the binomial theorem and rewrite the expression in terms of the power of $f(z)$ to get

$$f(z)^d + (p_{d-1}(z^2) - dz)d(z)^{d-1} + \cdots = 0. \tag{102.15}$$

Now consider the difference of the expressions in (102.12) and (102.15). This yields an algebraic relation for $f(z)$ of degree $\leq d-1$, and this must be identically 0 by the minimality of d. In particular,

$$p_{d-1}(z^2) - dz = p_{d-1}(z). \tag{102.16}$$

So, let

$$p_{d-1}(z) = \frac{A(z)}{B(z)} \tag{102.17}$$

for $\gcd(A(z), B(z)) = 1$ as polynomials. Thus, (102.16) can be written as

$$\frac{A(z^2)}{B(z^2)} - dz = \frac{A(z)}{B(z)}.$$

This is equivalent to

$$A(z)B(z^2) = A(z^2)B(z) - dzB(z)B(z^2). \tag{102.18}$$

Note that (102.18) implies that

$$B(z^2) \mid A(z^2)B(z)$$

and therefore

$$B(z^2) \mid B(z) \tag{102.19}$$

since $\gcd(A(z^2), B(z^2)) = 1$. Now, by comparing degrees, we conclude that $B(z) =: b$ is a nonzero constant polynomial (it's nonzero because it's a denominator in (102.17)).

Thus (102.18) can be rewritten as

$$A(z) = A(z^2) - bdz. \tag{102.20}$$

By comparing degrees again, we must have that $A(z) =: a$ is a constant. This forces $b = 0$ in order for (102.20) to be valid. This is a contradiction. Therefore, $f(z)$ is a transcendental function. □

Theorem 102-M: (Mahler, 1929) Let

$$f(z) := \sum_{n=0}^{\infty} z^{2^n}$$

for $|z| < 1$. Then, $f(\alpha)$ is transcendental for all algebraic $\alpha \neq 0$ with $|\alpha| < 1$.

Exercise: Show that

$$\sum_{n=1}^{\infty} \frac{z^{2^n}}{1 - z^{2^{n+1}}}$$

is actually an algebraic function. Use this to evaluate

$$\sum_{n=0}^{\infty} \frac{1}{F_{2^n}}$$

as an explicit algebraic number. F_n is the n^{th} Fibonacci number.

Chapter 15

Uniform Distribution

15.1 Basic Theorems

Recall: If $\frac{a}{b} \neq \frac{p}{q}$ are distinct rationals, then

$$0 \neq \left| \frac{a}{b} - \frac{p}{q} \right| = \frac{|aq - bp|}{bq} \geq \frac{1}{bq}.$$

Thus, if $\theta = \frac{a}{b} \in \mathbb{Q}$, then if $|q\theta - p| \neq 0$, then

$$|q\theta - p| \geq \frac{1}{b}$$

is bounded away from 0. Thus if $\theta \in \mathbb{R}$ and there exists an infinite sequence of integer pairs such that

$$0 \neq |q_n\theta - p_n| \to 0 \tag{103.1}$$

as $n \to \infty$, then θ is irrational.

Dirichlet's Theorem: Given $\theta \in \mathbb{R}$ and $N > 0$ arbitrarily large, there exists an integer q satisfying

$$\|q\theta\| < \frac{1}{N}, \quad 0 < q \leq N.$$

Recall that $\|q\theta\|$ is the distance between q and the nearest integer.

In particular, if θ is irrational, then there are infinitely many solutions to the inequality

$$\|q\theta\| < \frac{1}{q},$$

i.e.,

$$\left| \theta - \frac{p}{q} \right| < \frac{1}{q^2}.$$

Remark: We may interpret (103.1) as saying that the sequence of fractional parts $\{n\theta\}$ will have either 0 or 1 as a cluster. Kronecker extended this significantly.

Theorem 103: (Kronecker, 1884) If θ is irrational, then the sequence $\{n\theta\}$ of fractional parts of multiples of θ is dense in $[0, 1]$.

Remark: This property characterizes the irrationals because if $\theta \in \mathbb{Q}$, say $\theta = \frac{a}{b}$, then $\{n\theta\}$ is one of the values

$$\frac{0}{b}, \frac{1}{b}, \frac{2}{b}, \ldots, \frac{b-1}{b}$$

and so clustering can only occur around these points.

Equivalently, given $\epsilon > 0$ and $N > 0$ and any $\alpha \in [0, 1)$, there exist integers n and p such that

$$n\theta - p - \alpha| < \epsilon., \quad n > N. \tag{103.2}$$

Proof 1: Given irrational θ, by **Dirichlet's Theorem** there exists an integer $q > N$ such that

$$\begin{array}{c} 0 < \{q\theta\} < \epsilon, \qquad \text{or,} \\ 1 - \epsilon < \{q\theta\} < 1. \end{array} \tag{103.3}$$

Now consider the sequence

$$\{q\theta\}, \{2q\theta\}, \{3q\theta\}, \ldots.$$

Note that the sequence of fractional parts satisfies a linearity condition. In view of the linearity, the sequence in (103.4) provides a "mesh" in $[0, 1)$ with gaps no bigger than ϵ. Thus, there exists at least one member of the sequence that is of distance $< \epsilon$ from α, i.e., there exists $n = mq$ and p integers, such that

$$|n\theta - p - \alpha| < \epsilon$$

which is (103.2). □

Proof 2: This proof does not rely on **Dirichlet's Theorem**. If θ is irrational, then the sequence $\{n\theta\}$ consists of distinct numbers, all in $[0, 1)$. Thus there is at least one cluster point for this sequence in $[0, 1]$. If we denote $\{n\theta\}$ by P_n, then given $\epsilon > 0$, there are pairs P_n, P_{n+r} with r arbitrarily large such that the difference

$$|P_n - P_{n+r}| < \epsilon. \tag{103.5}$$

Let $|\overrightarrow{P_n P_{n+r}}|$ denote a vector from P_n to P_{n+r}, which could wrap around 1, in which case we define

$$|\overrightarrow{P_n P_{n+r}}| = |1 - P_n| + |P_{n+r}|. \tag{103.6}$$

By the linearity property of P_n, we can say that

$$|\overrightarrow{P_1 P_{1+r}}| < \epsilon. \tag{103.7}$$

Now, consider the sequence

$$\overrightarrow{P_1 P_{1+r}}, \overrightarrow{P_{1+r} P_{1+2r}}, \overrightarrow{P_{1+2r} P_{1+3r}}, \ldots \tag{103.8}$$

which yields a mesh in $[0, 1]$ with gaps $< \epsilon$. The proof is completed as in **Proof 1**. □

Proof 3: To avoid the discussion of crossings, we consider points on a circle of unit radius, i.e., we map

$$[0, 1) \longrightarrow \bigcirc.$$

Let $S = \big\{\{n\theta\}\big\}_{n=1}^{\infty}$. We are claiming that

$$\overline{S} = \bigodot. \tag{103.7}$$

Suppose not. Then, there exists α on the circle and an interval around α with null intersection with S, i.e., there exists $\delta, \delta' > 0$ such that

$$(\alpha - \delta, \alpha + \delta') \cap S \neq \emptyset. \tag{103.8}$$

From all such intervals, choose the maximal one and denote this by $I(\alpha)$. Again by linearity, we see that $I(\alpha)$ and $I(\alpha - \theta)$ are congruent, i.e.

$$I(\alpha) \equiv I(\alpha - \theta). \tag{103.9}$$

So, consider the sequence of intervals

$$I(\alpha), I(\alpha-\theta), I(\alpha-2\theta), \ldots. \tag{103.10}$$

So, two of these intervals must intersect. This immediately will yield an interval around α with a larger δ or δ' and null intersection with S, which contradicts the maximality of $I(\alpha)$. Thus, $\overline{S} = \bigodot$. $\square$

We have the following quantitative version of **Dirichlet's Theorem**:
Theorem: Given an irrational θ, there exist infinitely many pairs of integers (p, q) such that

$$0 \neq |q\theta - p| < \frac{1}{q}. \tag{104.1}$$

Is there a quantitative version of **Kronecker's Theorem** comparative to **Dirichlet's Theorem**? Yes.

Theorem 104: (Kronecker's Theorem - Quantitative Version) Let θ be irrational and let $\alpha \in \mathbb{R}$. Then for each $N > 0$ (arbitrarily large) there exist integers n and p such that

$$|n\theta - p - \alpha| < \frac{3}{n}, \;\; n > N. \tag{104.2}$$

Remark: It is not required that $\alpha \in (0, 1]$ since we may shift p by an integer if necessary.

Remark: Note that the inequality has an absolute constant 3 for all θ and α, which is bigger than the constant in **Dirichlet's Theorem**, but has the advantage of not depending on α.

Proof: Given N, by **Dirichlet's Theorem**, there exist integers q and r such that

$$0 \neq |q\theta - r| < \frac{1}{q}, \quad q > 2N. \tag{104.3}$$

Next, given α, define Q to be the integer nearest to $q\alpha$, i.e.,

$$|q\alpha - Q| \leq \frac{1}{2} \tag{104.4}$$

but there are two choices of Q is (104.4) has equality. Pick either one.

In (104.3), we may assume without loss of generality that $\gcd(r, q) = 1$. Thus there exist integers $u, v \in \mathbb{Z}$ such that

$$Q = vr - uq. \tag{104.5}$$

Since shifts $v \mapsto v + \lambda q$ and $u \mapsto u + \lambda r$ yield the same value in (104.5) with $\lambda \in \mathbb{Z}$, we may choose v such that

$$|v| \leq \frac{q}{2}. \tag{104.6}$$

With these choices,

$$q(v\theta - u - \alpha) = v(q\theta - r) - (q\alpha - Q). \tag{104.7}$$

So by (104.5), we have that

$$\begin{aligned} |q(v\theta - u - \alpha)| &\leq |v||q\theta - r| + |q\alpha - Q| \\ &< \frac{q}{2} \cdot \frac{1}{q} + \frac{1}{2} \\ &= \frac{1}{2} + \frac{1}{2} \\ &= 1, \end{aligned}$$

by (104.3) - (104.7). So,

$$|v\theta - u - \alpha| < \frac{1}{q}.$$

Now take

$$n = q + v \quad \text{and} \quad p = r + u. \tag{104.9}$$

Then,

$$N < \frac{q}{2} \leq n \leq q + \frac{q}{2} = \frac{3q}{2}. \tag{104.10}$$

Hence,

$$\begin{aligned} |n\theta - p - \alpha| &= |(q+v)\theta - (r+u) - \alpha| \\ &= |(v\theta - u - \alpha) + (q\theta - r)| \\ &\leq |v\theta - u - \alpha| + |q\theta - r| \\ &< \frac{1}{q} + \frac{1}{q} = \frac{2}{q} < \frac{3}{N}. \end{aligned}$$

Since this is true for all $n > N$, the proof is complete. □

15.2 An Application of Kronecker's Theorem to Geometry

The Problem of the Reflected Ray

Consider a square whose sides are reflecting mirrors. From a point inside the square, a ray of light emanates. What is the path of this light?

Theorem 105: (König & Szücs, 1913) The path is either periodic or is dense in the square. It is periodic if the slope is rational and dense when the slope is irrational.

Proof: For simplicity (and without loss of generality), we take the square to be the unit square, with the center of the square as the origin. Let the light emanate from a point $P = (a, b)$. Note that the set of images of P reflected over all mirrors repeatedly consists of

(A) $\{a + 2\ell, b + 2m\}$,

(B) $\{a + 2\ell, -b + 2m + 1\}$,

(C) $\{-a + 2\ell + 1, b + 2m\}$,

(D) $\{-a + 2\ell + 1, -b + 2m + 1\}$.

Observe that in each square there is precisely one image of P. Also, see that if the ray of light emanating from P is extended infinitely as a straight line, then the segments of this line in each square correspond bijectively with the segments between mirrors of the ray bounding around the original square. With λ, μ representing the direction cosines of the line, any point on the line is given by

$$x = a + \lambda t, \quad y = b + \mu t, \quad \text{slope} = \frac{\mu}{\lambda}, \quad t \in \mathbb{R}. \tag{105.1}$$

Note that the path of the ray of light is periodic if the line passes through an image of P of the type (A), i.e., there exists $t \in \mathbb{R}$ such that

$$\lambda t = 2\ell, \quad \mu t = 2m. \tag{105.2}$$

Therefore,

$$\frac{\mu}{\lambda} = \frac{m}{\ell} \in \mathbb{Q}.$$

This proves one part of the theorem.

Now let the slope be irrational. Pick any point (ξ, η) in the square. We want to show that the ray gets arbitrarily close to (ξ, η). This is equivalent to saying that the straight line of the ray is arbitrarily close to an image of (ξ, η) of type (A). So, we want to show that

$$|a + \lambda t - \xi - 2\ell| < \epsilon, \qquad |b + \mu t - \eta - 2m| < \epsilon, \tag{105.3}$$

for all $\epsilon > 0$. First, choose

$$t = \frac{\eta + 2m - b}{\mu}$$

which makes the second inequality in (105.3) zero and hence satisfies.

With this choice of t, we have

$$\begin{aligned} a + \lambda t - \xi - 2\ell &= a + \lambda \left(\frac{\eta + 2m - b}{\mu} \right) - \xi - 2\ell \\ &= 2m\frac{\lambda}{\mu} + (a - \xi) + \frac{\lambda(\nu - b)}{\mu} - 2\ell. \end{aligned} \tag{105.4}$$

Thus,

$$|a + \lambda t - \epsilon - 2\ell| < \epsilon$$

and hence

$$|m\theta - \omega - \ell| < \frac{\epsilon}{2}$$

where

$$\theta = \frac{\lambda}{\mu} \notin \mathbb{Q}, \qquad \omega = \frac{(b-\eta)\lambda}{2\mu} = \frac{a-\xi}{2} \in \mathbb{R}. \tag{105.5}$$

15.3 Simultaneous Approximation of Real Numbers

We are given n real numbers $\theta_1, \theta_2, \ldots, \theta_n$. Note that it is trivial to find integers $q_1, q_2, \ldots, q_n$ such that

$$\|q_1\theta_1\|, \|q_2\theta_2\|, \ldots, \|q_n\theta_n\|$$

are all small, because each q_i exists by **Dirichlet's Theorem**. This, the real interesting case is:

Problem: Determine a single integer q such that

$$\|q\theta_i\| < \epsilon \tag{106.1}$$

for some $\epsilon > 0$ and for all i. How effectively can this be done?

Theorem 106: (Dirichlet's Theorem in n Dimensions) Give n real numbers $\theta_1, \theta_2, \ldots, \theta_n$ and an integer N arbitrarily large, there exists an integer q satisfying the following:

$$0 < q < N^n \tag{106.2a}$$

and

$$\|q\theta_i\| < \frac{1}{N} \tag{106.2b}$$

Note that we can rewrite (106.1) as

$$\left|\theta_i - \frac{p_i}{q}\right| < \frac{\epsilon}{q} \tag{106.3a}$$

and we can rewrite (106.2b) as

$$\left|q\theta_i - \frac{p_i}{q}\right| < \frac{1}{qN}. \tag{106.3b}$$

It is important to see that we are not assuming that any of the θ_i are irrational.

Remark: Clearly, (106.2) implies (106.1) with $N > \dfrac{1}{\epsilon}$. Additionally, if θ_i is rational for some i, we do permit $\|q\theta_i\| = 0$, and so (106.2) could be trivially valid for this θ_i.

Proof: We establish this proof by using the Pigeon Hole Principal in n dimensions. Divide the n-dimensional cube $[0,1)^n$ into N^n subcubes, where each subcube has side length $\frac{1}{N}$. Now consider the $N^n + 1$ points in $[0,1)^n$ given by

$$(\{m\theta_1\}, \{m\theta_2\}, \ldots, \{m\theta_n\}) \tag{106.4}$$

where $m = 0, 1, \ldots, N^n$. Note we are using $\{r\}$ to mean the fractional part of r.

Now, by the Pigeon Hole Principal, at least two of these points lie in the same subcube, i.e., there exist integers $0 \le m < m' \le N^n$ such that

$$|\{m\theta_i\} - \{m'\theta_i\}| < \frac{1}{N} \tag{106.5}$$

for $i = 1, 2, \ldots, n$. Thus we set

$$q = m - m' \tag{106.6a}$$

and

$$p_i = [m'\theta_i] - [m\theta_i] \tag{106.6b}$$

and note that (106.2) is satisfied and $0 < q < N^n$. $\square$

Example: Consider $\theta_1 := \frac{4}{5}$ and $\theta_2 := \sqrt{2}$. Then, we can first make $\left|\sqrt{2} - \frac{p}{q}\right|$ small by **Dirichlet's Theorem**. Now we can approximate $\frac{4}{5}$ by

$$\left|\frac{4}{5} - \frac{4q}{5q}\right|$$

and subsequently actually approximate $\sqrt{2}$ by

$$\left|\sqrt{2} - \frac{5p}{5q}\right|.$$

We have the same denominator for both approximations, which is what we need. Granted, this approximation is not as strong as the case with $\sqrt{2}$ on its own, but this is the price paid for approximating multiple real numbers simultaneously.

Remark: If each θ_i is irrational, then $\|q\theta_i\| \neq 0$ for all i, and so we can make N larger to require a new q an infinite number of times. (If the norm could equal zero, then we could find a q where this happens and never need a new q.) So, as $N \to \infty$ in (106.2), we are guaranteed that infinitely many qs will be generated. Hence we arrive at the following:

Theorem 107: Let $\theta_1, \theta_2, \ldots, \theta_n$ be irrational. Let N be arbitrarily large. Then, there exists an integer q and integers $p_1, p_2, \ldots, p_n$ such that

$$0 < \left|\theta_i - \frac{p_i}{q}\right| < \frac{1}{qN} \tag{107.1}$$

for $i = 1, 2, \ldots, n$. In particular

$$0 < \left|\theta_i - \frac{p_i}{q}\right| < \frac{1}{q^{1+1/n}} \tag{107.2}$$

has infinitely many simultaneous solutions p_i/q. Setting $n = 1$ gives us the classical **Dirichlet Theorem**.

Remark: Recall that Kronecker took the fact from Dirichlet that the points $\|q\theta\|$ cluster near either 0 or 1, and then proved that in fact they cluster everywhere, i.e., they are dense in $[0, 1]$. Motivated by this, we will ask the same question on the n-dimensional cube. The above theorem tells us that the points $(\{q\theta_1\}, \ldots, \{q\theta_n\})$ cluster around at least one corner of $[0, 1)^n$. So we ask: When is $(\{q\theta_1\}, \ldots, \{q\theta_n\})_{q\in\mathbb{N}}$ dense in $[0, 1)^n$?

Observations: If any θ_i is rational and equals $\frac{a}{b}$, then $\{q\theta_i\}$ can take only the values

$$\frac{0}{b}, \ \frac{1}{b}, \ \ldots, \ \frac{b-1}{b}.$$

So clearly, we will not have a dense distribution in the case of θ_i rational. Thus it is necessary that the θ_i are all irrational. Suppose θ_1 and θ_2 satisfy a linear relation over $\mathbb{Q}$, such as

$$a\theta_1 + b\theta_2 = c \tag{108.1}$$

for $a, b, c \in \mathbb{Z}$ (they are actually rational, but we can clear denominators). Then

$$aq\theta_1 + bq\theta_2 = qc$$

for all $q \in \mathbb{Z}$, and so

$$a\{q\theta_1\} + b\{q\theta_2\} = qc - a[q\theta_1] - b[q\theta_2] =: d \in \mathbb{Z}. \tag{108.2}$$

Note that

$$|d| < |a| + |b| \tag{108.3}$$

since the fractional terms are less than 1. Thus, $a\{q\theta_1\} + b\{q\theta_2\}$ must be within $[0, 1)^2$ but on a finite set of lines segments

$$ax + by = d \tag{108.4}$$

and so the distribution cannot be dense. If we require that $1, \theta_1, \ldots, \theta_n$ are linearly independent over $\mathbb{Q}$, then this is actually sufficient to guarantee the n-dimensional version of **Kronecker's Theorem**.

Theorem 108: (Kronecker's Theorem in k Dimensions) Let $1, \theta_1, \theta_2, \ldots, \theta_k$ be linearly independent over $\mathbb{Q}$. Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be arbitrary real numbers and let $\epsilon > 0$ be arbitrarily small and $N > 0$ arbitrarily large be given. Then, there exists an integer q and integers $p_1, p_2, \ldots, p_k$ such that

$$|q\theta_i - p_i - \alpha_i| < \epsilon \tag{108.5}$$

for all $i = 1, 2, \ldots, k$ and $q > N$.

Remark: $1, \theta_1, \theta_2, \ldots, \theta_k$ linearly independent over $\mathbb{Q}$ implies that all θ_i are irrational. Additionally, we do not require that $\alpha_i \in [0, 1)$, since we may shift p_i by $[\alpha_i]$. Lastly, there is a version of this theorem where the hypotheses are weaker and the result is weaker, but the two are actually equivalent. This theorem is:

Theorem 108': Let $\theta_1, \theta_2, \ldots, \theta_k$ be linearly independent over $\mathbb{Q}$. (Note we have dropped that 1, so up to one of the θ_i may be rational.) Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be arbitrary real numbers, with $\epsilon > 0$ and $T > 0$ given. Then, there exists $t > T$ (not necessarily an integer) and integers $p_1, p_2, \ldots, p_k$ such that

$$|t\theta_i - p_i - \alpha_i| < \epsilon \tag{108.6}$$

and $t > T$.

Proof that Theorem 108' $\Longrightarrow$ Theorem 108: We assume without loss of generality that all $\theta_i \in [0, 1)$ and that $\epsilon < 1$. Given $\theta_1, \theta_2, \ldots, \theta_k$ and $\alpha_1, \alpha_2, \ldots, \alpha_k$ as in the hypothesis of **Theorem 108**, we apply **Theorem 108'** to the sets of $k + 1$ numbers

$$\theta_1, \theta_2, \ldots, \theta_k, 1 \qquad \text{and} \qquad \alpha_1, \alpha_2, \ldots, \alpha_k, 0 \tag{108.7}$$

and with $\epsilon/2$ in the place of ϵ in **Theorem 108'**. As per **Theorem 108**, the numbers, $1, \theta_1, \theta_2, \ldots, \theta_k$ are linearly independent, which means we are applying **Theorem 108'** to the $k + 1$ numbers $\theta_1, \theta_2, \ldots, \theta_k, 1$, which are linearly independent. Now, by **Theorem 108'**, one can find a real number

$$t > N + 1 \tag{108.8}$$

such that

$$\left.\begin{array}{c} |t\theta_i - p_i - \alpha_i| < \dfrac{\epsilon}{2}, \quad \text{for } i = 1, 2, \ldots, k \\ |t - p_{k+1} - 0| = |t - p_{k+1}| < \dfrac{\epsilon}{2} \end{array}\right\}. \tag{108.9}$$

Note that by (108.9) and (108.8), we have that

$$p_{k+1} > N. \tag{108.10}$$

This is the integer $q(:= p_{k+1})$ that will work in **Theorem 108**. That is, by (108.9), we have that

$$\begin{aligned} |p_{k+1}\theta_i - p_i - \alpha_i| &= |t\theta_i - p_i + \alpha_i + p_{k+1}\theta_i - t\theta_i| \\ &\leq |t\theta_i - p_i - \alpha_i| + |\theta_i||p_{k+1} - t| \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned} \tag{108.11}$$

for $i = 1, 2, \ldots, k$, which is the assertion of **Theorem 108**. □

Proof that Theorem 108 $\Longrightarrow$ Theorem 108': We make the preliminary observation that both forms of **Kronecker's Theorem** enjoy an "additive" property. To be more precise, if $\theta_1, \theta_2, \ldots, \theta_k$ are given satisfying the respective hypothesism, and if either theorem works for choices $\alpha_1, \alpha_2, \ldots, \alpha_k$ and $\beta_1, \beta_2, \ldots, \beta_k$

of reals, namely that the differences in (108.5) and (108.8) can be made arbitrarily small with the αs and βs respectively, and so can be made less than $\epsilon/2$, then the differences of the k real numbers $\alpha_1+\beta_1, \alpha_2+\beta_2, \cdots, \alpha_k+\beta_k$ can be made less than ϵ, because

$$|q\theta_i - p_i - \alpha_i| < \frac{\epsilon}{2}, \qquad |q'\theta_i - p_i' - \beta_i| < \frac{\epsilon}{2}, \quad i = 1, 2, \ldots, k \tag{108.12}$$

implies

$$|(q+q')\theta_i - (p_i+p_i') - (\alpha_i+\beta_i)| < \epsilon, \quad i = 1, 2, \ldots, k. \tag{108.13}$$

Similarly,

$$|t\theta_i - p_i - \alpha_i| < \frac{\epsilon}{2}, \qquad |t'\theta_i - p_i' - \beta_i| < \frac{\epsilon}{2}, \quad i = 1, 2, \ldots, k \tag{108.14}$$

implies

$$|(t+t')\theta_i - (p_i+p_i') - (\alpha_i+\beta_i)| < \epsilon, \quad i = 1, 2, \ldots, k. \tag{108.15}$$

We note in addition that both **Theorems 108 & 108'** are valid for $k = 1$.

Now consider $\theta_1, \theta_2, \ldots, \theta_{k+1}$ (with $k+1 \geq 2$), fitting the hypothesis of **Theorem 108'** which means that they are linearly independent over $\mathbb{Q}$. This means that the set

$$\frac{\theta_1}{\theta_{k+1}}, \frac{\theta_2}{\theta_{k+1}}, \ldots, \frac{\theta_k}{\theta_{k+1}}, 1 \tag{108.16}$$

are linearly independent over $\mathbb{Q}$. We now apply **Theorem 108** to the numbers in (108.16), given $\alpha_1, \alpha_2, \ldots, \alpha_k$ and N. Thus, there exists $q > N$ and integers $p_1, p_2, \ldots, p_k$ such that

$$\left| q\frac{\theta_i}{\theta_{k+1}} - p_i - \alpha_i \right| < \epsilon, \quad i = 1, 2, \ldots, k, \;\; q > N. \tag{108.17}$$

We may think of (108.17) as

$$|t\theta_i - p_i - \alpha_i| < \epsilon, \;\; i = 1, 2, \ldots, k \tag{108.18}$$

with $t = \dfrac{q}{\theta_{k+1}} > N$.

Note that trivially, due to (108.18), we have that

$$|t\theta_{k+1} - q| = 0 < \epsilon \tag{108.19}$$

and so $q =: p_{k+1}$ is the choice p_{k+1}. Thus (108.18) and (108.19) show that **Theorem 108'** is valid for the case

$$\theta_1, \theta_2, \ldots, \theta_k, \theta_{k+1}, \qquad \text{and} \qquad \alpha_1, \alpha_2, \ldots, \alpha_k, 0. \tag{108.20}$$

Simiarly, **Theorem 108'** is valid for the case

$$\theta_1, \theta_2, \ldots, \theta_k, \theta_{k+1}, \qquad \text{and} \qquad 0, 0, \ldots, 0, \alpha_{k+1}. \tag{108.21}$$

So from (108.20) and (108.21) we see by the additive property that **Theorem 108'** holds for $\theta_1, \theta_2, \ldots, \theta_{k+1}$ and $\alpha_1, \alpha_2, \ldots, \alpha_{k+1}$ arbitrary. □

15.4 Kronecker's Theorem in Two Dimensions

We will prove the first version of **Kronecker's Theorem**, namely **Theorem 108** in the case $k = 2$. This then can be extended to higher dimensions.

Proof: (Lettenmeyer) Lettenmeyer's proof given here with the case $k = 1$ was **Proof (ii)** of **Kronecker's Theorem** given above.

So, let θ_1, θ_2 be reals such that $1, \theta_1, \theta_2$ are linearly independent over $\mathbb{Q}$. In particular, both θ_1, θ_2 are irrationals. Without loss of generality, we assume

$$0 < \theta_1, \theta_2 < 1. \tag{108.22}$$

Consider the points P_n in the unit square whose coordinates are

$$P_n = (\{n\theta_1\}, \{n\theta_2\}). \tag{108.23}$$

We need to show that the set of points P_n is dense in $[0,1) \times [0,1)$. Note that the P_n are all distinct points and no P_n lies on the side of the square because $\{n\theta_i\}$ cannot be 0 or 1. As before, we construct the directed line segment connecting P_n and P_{n+r} as a vector and write $\overrightarrow{P_nP_{n+r}}$ for this vector.

By linearity, we note that given $\overrightarrow{P_nP_{n+r}}$, if we take any P_m and draw a line segment from P_m equal in length and parallel to $\overrightarrow{P_nP_{n+r}}$ to get an end point Q, then Q is indeed P_{m+r}, with the same convention being adopted.

Since P_n is an infinite sequence of distinct points in the unit square, the sequence has a cluster point. Thus for each $\epsilon > 0$, there are vectors $\overrightarrow{P_nP_{n+r}}$ with r arbitrarily large and with

$$|\overrightarrow{P_nP_{n+r}}| < \epsilon. \tag{108.24}$$

To avoid anomalies, we choose

$$\epsilon < \min\{\theta_1, 1-\theta_1, \theta_2, 1-\theta_2\}. \tag{108.25}$$

Thus, any vector of length less than ϵ emanating from P_1 will not cross any side of the square. There are now two cases to consider:

Case 1: There are two vectors of length less than ϵ which are not parallel.

In this case, by linearity, we may make these vectors emanate from P_1 and draw a lattice with these two vectors with the convention as before. This provides a mesh within $[0,1) \times [0,1)$ such that every point in the unit square is within ϵ distance of an end point of this mesh. That proves the theorem in this case.

Case 2: All vectors of length less than ϵ are parallel.

By linearity all ϵ-vectors emanating from P_1 will be parallel. This means there are integers r and s arbitrarily large such that the points, P_1, P_r, P_s are collinear. This means

$$0 = \det \begin{pmatrix} \theta_1 & \theta_2 & 1 \\ \{r\theta_1\} & \{r\theta_2\} & 1 \\ \{s\theta_1\} & \{s\theta_2\} & 1 \end{pmatrix} = \det \begin{pmatrix} \theta_1 & \theta_2 & 1 \\ r\theta_1 - [r\theta_1] & r\theta_2 - [r\theta_2] & 1 \\ s\theta_1 - [s\theta_1] & s\theta_2 - [s\theta_2] & 1 \end{pmatrix}. \tag{108.25}$$

Now, (108.25) implies that

$$\det \begin{pmatrix} \theta_1 & \theta_2 & 1 \\ [r\theta_1] & [r\theta_2] & r-1 \\ [s\theta_1] & [s\theta_2] & s-1 \end{pmatrix} = 0. \tag{108.26}$$

By expanding this determinant, we may write it as

$$a\theta_1 + b\theta_2 + c = 0 \tag{108.27}$$

where $a, b, c \in \mathbb{Z}$ with

$$a = \det \begin{pmatrix} [r\theta_2] & r-1 \\ [s\theta_2] & s-1 \end{pmatrix}, \quad b = -\det \begin{pmatrix} [r\theta_1] & r-1 \\ [s\theta_1] & s-1 \end{pmatrix}. \tag{108.28}$$

Since $1, \theta_1, \theta_2$ are linearly independent, we must have

$$a = b = c = 0. \tag{108.29}$$

In particular,

$$\det \begin{pmatrix} [r\theta_2] & r-1 \\ [s\theta_2] & s-1 \end{pmatrix} = 0 = [r\theta_2](s-1) - [s\theta_2](r-1), \tag{108.30}$$

which is the same as

$$\frac{[s\theta_2]}{s-1} = \frac{[r\theta_2]}{r-1} \tag{108.31}$$

with r and s arbitrarily large. If we keep s fixed and let r alone tend to ∞, we get

$$\lim_{r\to\infty} \frac{[r\theta_2]}{r-1} = \lim_{r\to\infty} \frac{r\theta_2 - \{r\theta_2\}}{r-1} = \theta_2 \tag{108.32}$$

which is irrational, forcing

$$\frac{[s\theta_2]}{s-1} = \lim_{r\to\infty} \frac{[r\theta_2]}{r-1} = \theta_2 \tag{108.33}$$

to be irrational, which is impossible because the left-hand side of (108.33) is rational. Thus, this case cannot hold, and so only case 1 holds, which means **Theorem 108** for $k = 2$ is true. □

Remark: Lettenmeyer's proof for $k = 2$ (and $k = 1$) extends to all $k \geq 2$ by considering linearly independent/dependent ϵ-vectors and $(k+1) \times (k+1)$ determinants. But what we will do now is to deduce the result for all k by induction on k. This proof is due to Estermann.

Proof of Kronecker's Theorem in Multiple Dimensions: (Estermann)

For a certain $k \geq 2$, we assume that the theorem is true for dimension $k-1$. For the purpose of induction, it is useful to note that one can actually establish a stronger version of the theorem, namely: If $1, \theta_1, \theta_2, \ldots, \theta_k$ are linearly independent and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are arbitrary real numbers, then given any $\epsilon > 0$ (arbitrarily small), and $\omega > 0$ (arbitrarily large) and $\lambda \neq 0$, there exist integers $n, p_1, p_2, \ldots, p_k$ such that

$$|n\theta_i - p_i - \alpha_i| < \epsilon \quad i = 1, 2, \ldots, k, \quad \operatorname{sgn}(n) = \operatorname{sgn}(\lambda), \quad |n| > \omega. \tag{108.34}$$

For the induction to work, we will have to show that the truth of (108.34) for $k-1$ implies its truth for k¡ and that (108.34) holds for $k = 1$.

For the case of dimension k, first note that **Dirichlet's Theorem** in k dimensions (**Theorem 106**) guarantees that there exists an integer $q > 0$ and integers, $b_1, b_2, \ldots, b_k$ such that

$$|q\theta_i - b_i| < \frac{\epsilon}{2}, \quad i = 1, 2, \ldots, k. \tag{108.35}$$

Since θ_i are irrational, each $q\theta_i - b_i \neq 0$. Thus we consider the numbers

$$\varphi_i = \frac{s\theta_i - b_i}{s\theta_k - b_k}, \quad i = 1, 2, \ldots, k \tag{108.36}$$

of which $\varphi_k = 1$. Note that $\varphi_1, \varphi_2, \ldots, \varphi_k$ are linearly independent, because a linear dependence between them would imply one for $\theta_1, \theta_2, \ldots, \theta_k$, which cannot be. Thus we apply the theorem in the stronger form (108.34) to

$$\varphi_1, \varphi_2, \ldots, \varphi_{k-1}, \varphi_k = 1 \tag{108.37}$$

which is the k-dimensional case, with the choices

$$\begin{gathered} \beta_i = \alpha_i - \alpha_k \varphi_k, \quad i = 1, 2, \ldots, k-1 \\ \epsilon/2 \text{ in place of } \epsilon \\ \lambda(q\theta_k - b_k) \text{ in place of } \lambda \\ \Omega = (\omega + 1)|q\theta_k - b_k| + |\alpha_k| \text{ in place of } \omega. \end{gathered} \tag{108.38}$$

Thus by the induction hypothesis, there are integers $c_1, c_2, \ldots, c_k$ such that

$$|c_k \varphi_i - c_i - \beta_i| < \frac{\epsilon}{2}, \quad i = 1, 2, \ldots, k-1, \quad |c_k| > \Omega, \quad \operatorname{sgn}(c_k) = \operatorname{sgn}(\lambda(q\theta_k - b_k)). \tag{108.39}$$

Substituting the values φ_i as in (108.36) and β_i as in (108.38) into (108.39), we may rewrite it as

$$\left| \frac{(c_k + \alpha_k)}{q\theta_k - b_k}(q\theta_i - b_i) - c_i - \alpha_i \right| < \frac{\epsilon}{2}, \quad i = 1, 2, \ldots, k, \tag{108.40}$$

where we have included $i = k$ in (108.40) because in that case, the inequality is trivially true because the left-hand side of (108.40) is zero.

Note that when $k = 1$, (108.40) is trivially true, and (108.39) holds as well when $k - 1 = 1$.

Now choose an integer N such that

$$\left| N - \frac{c_k + \alpha_k}{q\theta_k - b_k} \right| < 1. \tag{108.41}$$

Set $n := Nq$ and $p_i := Nb_i + c_i$. Then, we have that

$$\begin{aligned} |n\theta_i - p_i - \alpha_i| &= |N(q\theta_i - \beta_i) - c_i - \alpha_i| \\ &\leq \left| \frac{c_k + \alpha_k}{q\theta_k - b_k}(q\theta_i - b_i) - c_i - \alpha_i \right| + |q\theta_i - b_i| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \quad i = 1, 2, \ldots, k, \end{aligned} \tag{108.42}$$

by virtue of (108.39) and (108.40).

Note that by (108.38) and (108.39),

$$\frac{c_k + \alpha_k}{q\theta_k - b_k} \geq \frac{|c_k| - |\alpha_k|}{|q\theta_k - b_k|} > \frac{\Omega - |\alpha_k|}{|q\theta_k - b_k|} = \omega + 1 \tag{108.43}$$

and so by (108.41) and (108.43)

$$|N| > \omega \qquad \text{and} \qquad |n| = |N|q \geq N > \omega. \tag{108.44}$$

Also,

$$\operatorname{sgn}(n) = \operatorname{sgn}(N) = \operatorname{sgn}\left(\frac{c_k}{q\theta_k - b_k} \right) = \operatorname{sgn}(\lambda),$$

which means that (108.34) has been proved for the k-dimensional case. Thus, **Theorem 108** in the strong form has been established by induction. □

15.5 Uniform Distribution Modulo 1

The theory of uniform distribution was ushered in by Hermann Weyl in a class paper "Über die Gleichverteilung von Zahlen modulo Eins", published in Math. Ann. 77 (1916), pgs. 313-352. A good reference of this material is "Uniform Distribution of Sequences" by L. Kuipers and H. Niederreiter, published in 1976.

Definition: Given a sequence $\omega = \{x_n\}_{n\in\mathbb{N}}$ of real numbers and $E \subseteq [0,1)$, consider the counting function

$$A(E,N,\omega) = \sum_{\substack{n\geq N \\ \{x_n\}\in E}} 1. \tag{109.1}$$

We will focus on $E = [a,b)$, i.e., intervals. We say that ω is uniformly distributed modulo 1 if for every $[a,b) \subseteq [0,1)$ we have

$$\lim_{N\to\infty} \frac{A([a,b),N,\omega)}{N} = b-a. \tag{109.2}$$

Remark: Since

$$[a,b) = [0,b) \setminus [0,a) \tag{109.3}$$

and since

$$A([a,b),N,\omega) = A([0,b),N,\omega) - A([0,a),N,\omega) \tag{109.4}$$

it suffices to define "uniformly distributed" by

$$\lim_{N\to\infty} \frac{A([0,a),N,\omega)}{N} = a. \tag{109.5}$$

Definition: Define the characteristic function of an interval by

$$\chi_{[a,b)}(x) := \begin{cases} 1, & x\in[a,b) \\ 0, & \text{otherwise} \end{cases}. \tag{109.6}$$

Remark: Now note that

$$b-a = \int_0^1 \chi_{[a,b)}(x)\,dx, \tag{109.7}$$

and so (109.2) can be rewritten as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{1\leq n\leq N} \chi_{[a,b)}(x_n) = \int_0^1 \chi_{[a,b)}(x)\,dx. \tag{109.8}$$

This reformulation helps us extend (109.8) to step functions, and then to continuous functions. More precisely, we formulate the following theorem.

Theorem 110: A sequence $\omega = \{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_0^1 f(x)\,dx \tag{110.1}$$

for all continuous functions on $[0,1]$.

Proof: Let $\omega = \{x_n\}_{n\in\mathbb{N}}$ be uniformly distributed modulo 1. Then by (109.8), we see by linearity that (110.1) holds for all step functions

$$\psi(x) = \sum_{i=1}^{k} c_i \chi_{[a_i, a_{i+1})}(x) \tag{110.2}$$

where $0 = a_0 < a_1 < \cdots < a_{k+1} = 1$.

If $f \in \mathscr{C}([0,1])$ and $\epsilon > 0$ are given, then there exist two step functions ψ_1, ψ_2 satisfying

$$\psi_1(x) \le f(x) \le \psi_2(x) \tag{110.3}$$

and

$$0 < \int (\psi_2(x) - \psi_1(x))\, dx < \epsilon. \tag{110.4}$$

Therefore,

$$\begin{aligned}
\int_0^1 f(x)\, dx - \epsilon \le_0^1 \psi_1(x)\, dx & \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \psi_1(x_n) \\
&\le \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) \\
&\le \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) \\
&\le \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \psi_2(x_n) \\
&= \int_0^1 \psi_2(x)\, dx \\
&\le \int_0^1 f(x)\, dx + \epsilon.
\end{aligned} \tag{110.5}$$

We may let $\epsilon \to 0$ in (110.5) and conclude that (110.1) holds for f.

Conversely, assume that (110.1) holds for all $f \in \mathscr{C}([0,1])$. Let $\chi_{[a,b)}$ be given. Given $\epsilon > 0$, we may choose continuous functions f_1, f_2 such that

$$f_1(x) \le \chi_{[a,b)} \le f_2(x) \tag{110.6}$$

and

$$0 < \int (f_2(x) - f_1(x))\, dx < \epsilon. \tag{110.7}$$

Thus,

$$\begin{aligned}(b-a)-\epsilon &< \int_0^1 f_1(x)\,dx \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f_1(x_n) \\ &\le \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\chi(x_n) \\ &\le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\chi(x_n) \\ &\le \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f_2(x_n) \\ &= \int_0^1 f_2(x)\,dx \\ &\le (b-a)+\epsilon.\end{aligned}$$

We may now let $\epsilon \to 0$ to conclude that (110.1) holds for $\chi_{[a,b)}$. Then, $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. □

Corollary 110-1: $\omega := \{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if (110.1) holds for all Riemann integrable functions on $[0,1)$.

Remark: Note that in the above corollary we are assuming that the Riemann integrable functions are bounded. If they are unbounded functions (which are still Riemann integrable), then the improper integral is defined in the normal way on bounded functions, by

$$\int_a^b f(x)\,dx = \lim_{\substack{d\nearrow b\\ c\searrow a}}\int_c^d f(x)\,dx, \tag{110.9}$$

where f is necessarily bounded in the compact interval $[c,d]$.

Corollary 110-2: Let $\omega := \{x_n\}_{n\in\mathbb{N}}$ be a real sequences. Then, ω is uniformly distributed modulo 1 if and only if (110.1) holds for all real continuous functions on $\mathbb{R}$ which are periodic of period 1.

Proof: Saying that f is periodic of period 1 is equivalent to

$$f(x) = f(\{x\}). \tag{110.10}$$

We may then approximate any continuous f on $[0,1]$ by a periodic continuous function g such that

$$\int_0^1 ((f(x)-g(x))dx < \epsilon.\ \square \tag{110.11}$$

Corollary 110-3: $\omega := \{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if (110.1) holds for all complex valued functions on $\mathbb{R}$ which are periodic of period 1.

15.6 The Weyl Criterion

Theorem 111: The sequence $\omega := \{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} = 0 \tag{111.1}$$

for every $h \in \mathbb{Z} \setminus \{0\}$.

Remark: For $h = 0$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i (0) x_n} = 1 = \int_0^1 e^{2\pi i (0) x}\, dx = \int_0^1 1\, dx.$$

Therefore, (110.1) is true in the case $h = 0$.

Proof: Let $h \neq 0$ and let $\omega = \{x_n\}_{n\in\mathbb{N}}$ be uniformly distributed modulo 1. In this case, by one of the corollaries to **Theorem 110**, we have that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} = \int_0^1 e^{2\pi i h x}\, dx = \frac{e^{2\pi i h x}}{2\pi i h}\bigg|_0^1 = 0. \tag{111.2}$$

Therefore, (110.1) holds for all functions of the form $f_h(x) = e^{2\pi i h x}$, for $h \in \mathbb{Z}$. Additionally, this means that (110.1) is true for all linear combinations of $f_h(x)$, which is precisely the set of trigonometric polynomials.

Given any continuous periodic function f of period 1, by the **Weierstrass Approximation Theorem**, we know that f can be approximated by trigonometric polynomials. So, given $\epsilon > 0$, there exists a trigonometric polynomial $\psi(x)$ such that

$$\sup_{0\leq x\leq 1} |f(x) - \psi(x)| < \frac{\epsilon}{2}. \tag{111.3}$$

Therefore,

$$\begin{aligned}\left|\left(\frac{1}{N}\sum_{n=1}^{N} f(x_n)\right) - \int_0^1 f(x)\, dx\right| &\leq \frac{1}{N}\left|\sum_{n=1}^{N}(f(x_n) - \psi(x_n))\right| + \left|\left(\frac{1}{N}\sum_{n=1}^{N}\psi(x_n)\right) - \int_0^1 \psi(x)\, dx\right| \\ &\quad + \left|\int_0^1 (\psi(x) - f(x))dx\right| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}.\end{aligned}$$

for $N \geq N(\epsilon)$, for the middle term. (The first and last terms on the right in (111.4) are each $< \frac{\epsilon}{3}$ by virtue of (111.3).) Since this is true for every $\epsilon > 0$, we let $\epsilon \to 0$ and $N = N(\epsilon) \to \infty$ to conclude that (110.1) is valid for all periodic f of period 1. Hence ω is uniformly distibuted modulo 1. □

Theorem 112: (Consequence of Theorem 111) Let θ be irrational. Then, $\{n\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Proof: We need to check that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h n \theta} = 0 \tag{112.1}$$

for all $0 \in h \in \mathbb{Z}$. But then

$$\frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h n \theta} = \frac{e^{2\pi i h \theta}}{N} \cdot \frac{e^{2\pi h N \theta} - 1}{e^{2\pi i h \theta} - 1}. \tag{112.2}$$

Observe that if $h \neq 0$, then

$$e^{2\pi i h\theta} - 1 \neq 0.$$

Thus the right-hand side of (112.2) is

$$O_h\left(\frac{1}{n}\right) \to 0$$

as $N \to \infty$. Hence by the **Weyl Criterion**, we have that **Theorem 112** follows. □

Remark: Using the **Weyl Criterion**, one can establish the uniform distribution of other sequences. A particularly deep example is

$$\omega := \{p\theta\}$$

for a prime p. This uses a method of trigonometric sums due to I. M. Vinogradov.

15.7 Similar results on $\mathbb{Q}$

Remark: Clearly, **Theorem 112** characterizes the irrationals. A natural question is whether we can find similar results for the rationals. Consider $\mathbb{Q} \cap [0,1)$ ordered lexicographically as such:

$$\frac{0}{1}, \frac{0}{2}, \frac{1}{2}, \frac{0}{3}, \frac{1}{3}, \frac{2}{3}, \frac{0}{4}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \dots \tag{113.1}$$

where the m^{th} block is

$$\frac{0}{m}, \frac{1}{m}, \dots, \frac{m-1}{m}.$$

For now, we keep the repetitions.

Theorem 113: The sequence of rationals lexicographically ordered as in (113.1) is uniformly distributed modulo 1.

Proof: First observe that the number of members of the sequence ω up to the m^{th} block is

$$1 + 2 + \cdots + m = \frac{m(m+1)}{2} = T_m. \tag{113.2}$$

Recall that T_m is the m^{th} triangular number.

Given $N \in \mathbb{Z}$ large, determine m such that

$$T_m \leq N < T_{m+1}. \tag{113.3}$$

Now,

$$0 \leq N - T_m < O(\sqrt{n}). \tag{113.4}$$

The number of members of the sequence in the k^{th} block that lie within $[a,b)$ is exactly

$$\sum_{a \leq j/k < b} 1 = \sum_{ka \leq j < kb} 1 = k(b-a) + O(1). \tag{113.5}$$

Thus,

$$A([a,b), T_m, \omega) = \sum_{k=1}^{m} (k(b-a) + O(1)) = (b-a)T_m + O(m). \tag{113.6}$$

But then,

$$\begin{aligned} A([a,b), N, \omega) &= A([a,b), T_m, \omega) + (A([a,b), N, \omega) - A([a,b), T_m, \omega)) \\ &= (b-a)T_m + O(m) + O(m) \\ &= (b-a)N + O(\sqrt{N}). \end{aligned} \tag{113.7}$$

Therefore,

$$\frac{A([a,b), N, \omega)}{N} = (b-a) + O\left(\frac{1}{\sqrt{N}}\right). \tag{113.8}$$

This tends to $b-a$ as $N \to \infty$. Therefore, ω is uniformly distributed modulo 1. $\square$

Remark: Clearly, the case of real interest is when we remove duplicates from the sequence of rationals. We want to lexicographically order the rationals ignoring duplications:

$$\frac{0}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \dots. \tag{114.1}$$

This sequence will be called $\mathscr{F}$, after Farey. (Note that this sequence does not have the same ordering as the Farey fractions we studied earlier.) The number of fractions in the m^{th} block is $\varphi(m)$, where φ is the Euler function. So, the number of terms of $\mathscr{F}$ up to the m^{th} block is

$$\Phi(m) = \sum_{n=1}^{m} \varphi(n). \tag{114.2}$$

Lemma 114.1:

$$\Phi(m) = \frac{3m^2}{\pi^2} + O(m\log(m)). \tag{114.3}$$

Proof: The Euler function satisfies the fundamental relation

$$\sum_{d|n} \varphi(d) = n \tag{114.4}$$

and so by Möbius inversion, we have that

$$\varphi(n) = \sum_{d|n} \mu(d)\frac{n}{d}. \tag{114.4}$$

Therefore,

$$\begin{aligned}
\Phi(m) &= \sum_{n=1}^{m} \varphi(m) \\
&= \sum_{n=1}^{m} \sum_{d|n} \mu(d)\frac{n}{d} \\
&= \sum_{d\le m} \left(\mu(d) \sum_{k\le \frac{m}{d}} k \right) \\
&= \sum_{d\le m} \left(\mu(d) \left(\frac{\left[\frac{m}{d}\right]\left[\frac{m}{d}\right]+1}{2} \right) \right) \\
&= \sum_{d\le m} \left(\mu(d) \left(\frac{\frac{m^2}{d^2} + O\left(\frac{m}{d}\right)}{2} \right) \right) \\
&= \frac{m^2}{2} \sum_{d\le m} \left(\frac{\mu(d)}{d^2} + O\left(\sum_{d\le m} \frac{m}{d} \right) \right) \\
&= \frac{m^2}{2} \left[\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} - \sum_{d=m+1}^{\infty} \frac{\mu(d)}{d^2} \right] + O(m\log(m)) \\
&= \frac{3}{\pi^2} m^2 + O(m\log(m)),
\end{aligned} \tag{114.6}$$

because

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \frac{1}{\zeta(s)}, \quad \Re(s) > 1 \tag{114.7}$$

and we know

$$\zeta(2) = \frac{\pi^2}{6}. \tag{114.8}$$

Additionally,

$$\begin{aligned}\left|\sum_{d=m+1}^{\infty}\frac{\mu(d)}{d^2}\right| &\le \sum_{d=m+1}\frac{1}{d^2}\\ &= O\left(\int_{m+1}^{\infty}\frac{1}{t^2}\,dt\right)\\ &= O\left(\frac{1}{m}\right).\end{aligned} \tag{114.9}$$

We need a generalization of the Euler function. Define

$$\varphi(n,x) := \sum_{\substack{k\le x\\ \gcd(k,n)+1}} 1. \tag{114.10}$$

Note that $\varphi(n,n)=n$.

Lemma 114.2: For all $x>0$,

$$\varphi(n,x)=\frac{x}{n}\varphi(n)+O(d(n))$$

where $d(n)$ is the number of divisors of n.

Proof: By either Möbius Inversion or by the Inclusion-Exclusion Principle, we have that

$$\begin{aligned}\varphi(n,x) &= \sum_{m\le x}\sum_{d\mid \gcd(m,n)}\mu(d)\\ &= \sum_{d|n}\left(\mu(d)\sum_{\substack{m\le x\\ m|d}}1\right)\\ &= \sum_{d|n}\mu d\left[\frac{x}{d}\right]\\ &= \sum_{d|n}\mu d\frac{x}{d}-\sum_{d|n}\mu(d)\left\{\frac{x}{d}\right\}\\ &= x\frac{\varphi(n)}{n}+O(d(n)).\ \square\end{aligned} \tag{114.11}$$

Remark: The function $\varphi(n,x)$ is the fundamental counting function in Sieve questions. For example, a typical Sieve question is: Given a set A and a set P of primes p, and y large, estimate

$$S(A,P,y)=\sum_{\substack{n\in A\\ n\le x\\ \gcd(n,P_y)=1}}1$$

where

$$P_y=\prod_{\substack{p\in P\\ p\le y}}p.$$

Theorem 114: The sequence of fractions $\mathscr{F}$ is uniformly distributed in [0,1).

Proof: Given a large N, we determine m such that

$$\Phi(m)\le N<\Phi(m+1). \tag{114.12}$$

Thus, analogous to (113.3), we have

$$\begin{aligned} 0 &\le N - \Phi(m) \\ &< \Phi(m+1) - \Phi(m) = \varphi(m+1) \\ &\le m \\ &= O(\sqrt{N}). \end{aligned} \tag{114.13}$$

Next, given $0 \le \alpha < \beta < 1$, we consider the number of fractions in the k^{th} block that lie in $[\alpha, \beta)$. This is given by

$$\begin{aligned} \sum_{\substack{\alpha \le \frac{j}{k} < \beta \\ \gcd(j,k)=1}} 1 &= \sum_{\substack{\alpha k \le j < \beta k \\ \gcd(j,k)=1}} 1 \\ &= \varphi(k, \beta k) - \varphi(k, \alpha k) + O(1) \\ &= \beta\varphi(k) - \alpha\varphi(k) + O(d(k)). \end{aligned} \tag{114.14}$$

Thus,

$$\begin{aligned} A([\alpha,\beta), \Phi(m), \mathscr{F}) &= (\beta-\alpha)\sum_{k=1}^{m} \varphi(k) + O\left(\sum_{k=1}^{m} d(k)\right) \\ &= (\beta-\alpha)\Phi(m) + O(m\log(m)). \end{aligned} \tag{114.15}$$

Now,

$$\begin{aligned} A([\alpha,\beta), N, \mathscr{F}) &= A([\alpha,\beta), \Phi(m), \mathscr{F}) + A([\alpha,\beta), N, \mathscr{F}) - A([\alpha,\beta), \Phi(m), \mathscr{F}) \\ &= (\beta-\alpha)\Phi(m) + O(m\log(m)) + O(m) \\ &= (\beta-\alpha)N + O(\sqrt{N}\log(N)). \end{aligned} \tag{114.16}$$

Thus,

$$\frac{A([\alpha,\beta), N, \mathscr{F})}{N} = (\beta-\alpha) + O\left(\frac{\log(N)}{\sqrt{N}}\right) \tag{114.17}$$

and this goes to $\beta - \alpha$ as $N \to \infty$. Therefore, $\mathscr{F}$ is uniformly distributed in $[0, 1)$. □

Remark: The sum of $d(k)$ is

$$\sum_{k=1}^{m} d(k) = m\log(m) + (2\gamma - 1)m + O(\sqrt{m}). \tag{114.18}$$

This can be proved by the "hyperbola method". For our purpose in (114.15), the simpler estimate

$$\begin{aligned} \sum_{k \le m} d(k) &= \sum_{k \le m} \sum_{e \mid k} 1 \\ &= \sum_{e \le m} 1 \sum_{\substack{k \le m \\ k \mid e}} \\ &\le \sum_{e \le m} \frac{m}{e} \\ &= m \sum_{e \le m} \frac{1}{e} \\ &= O(m\log(m)). \end{aligned}$$

Theorem 115: Let $\rho = \{\rho_n\}$ be the rationals in $[0,1)$, written lexicographically with repetition allowed. Then, for every $0 \neq h \in \mathbb{Z}$, we have that

$$\sum_{n=1}^{N} e^{2\pi i h \rho_n} = O\left(\sigma(h) + \frac{\sqrt{N}}{h}\right) \tag{115.1}$$

for $N > h^2$.

Proof: Given N large, as before choose M such that

$$T_m \leq N < T_{m+1}. \tag{115.2}$$

Consider the Weyl sum over the k^{th} block:

$$\sum_{j=0}^{k} e^{2\pi i h \frac{j}{k}} = \begin{cases} 0, & k \nmid h \\ k, & k \mid h \end{cases}. \tag{115.3}$$

Thus the sum in (115.1) yields

$$\begin{aligned} \sum_{n=1}^{N} e^{2\pi i h \rho_n} &= \sum_{n=1}^{T_m} e^{2\pi i h \rho_n} + \sum_{n=T_m+1}^{N} e^{2\pi i h \rho_n} \\ &=: \Sigma_1 + \Sigma_2. \end{aligned} \tag{115.4}$$

Now,

$$\Sigma_1 = \sum_{k=1}^{m} \sum_{j=0}^{k-1} e^{2\pi i h \frac{j}{k}} = \sum_{\substack{k \leq m \\ k \mid h}} k \leq O(\sigma(h)). \tag{115.5}$$

Regarding Σ_2, we have

$$\Sigma_2 = \sum_{j=0}^{N-T_m} e^{2\pi i h \frac{j}{m+1}} = \frac{\zeta^{h\nu} - 1}{\zeta^h - 1} = O\left(\frac{1}{\zeta^h - 1}\right). \tag{115.6}$$

where

$$\zeta = e^{2\pi i/(m+1)}$$

and

$$\nu = N - T_m + 1.$$

At this stage, we get that

$$|e^z - 1| >> |z|$$

for $z \neq 0$ and $z \in \mathbb{C}$. Thus,

$$\Sigma_2 = O\left(\frac{m+1}{h}\right) = P\left(\frac{\sqrt{N}}{h}\right).$$

This completes the theorem. □

Theorem 116: Let $\mathscr{F} = \{f_n\}$ be the sequence of all reduces rationals in lexicographic order as before. Then, for any nonzero $h \in \mathbb{Z}$, we have

$$S_N := \sum_{n=1}^{N} e^{2\pi i h f_n} = O(\sqrt{N} d(h)) \tag{116.1}$$

where $d(h)$ is the divisor function. Consequently, by the **Weyl Criterion**, we have that $\{f_n\}$ is uniformly distributed in $[0,1)$.

Remark: In deriving an estimate for S_N, the role of the "Ramanujan Sums" will be crucial

Proof: Given large N, we determined the integer m for which

$$\Phi(m) \le N < \Phi(m+1), \tag{116.2}$$

where $\Phi(m) = \sum_{k\le m} \varphi(k)$, the sum of the Euler function. With m as above, we may write

$$S_N = \sum_{n\le \Phi(m)} e^{2\pi i h f_n} + \sum_{\Phi(m)+1\le n\le N} e^{2\pi i h f_n} =: \Sigma_1 + \Sigma_2. \tag{116.3}$$

Clearly,

$$|\Sigma_2| \le N - \Phi(m) < \Phi(m+1) < m = O(\sqrt{N}) \tag{116.4}$$

by **Lemma 114-1**. Now, write Σ_1 as

$$\begin{aligned}\Sigma_1 &= \sum_{k=1}^{m} \sum_{\substack{1\le j\le k\\ \gcd(j,k)=1}} e^{2\pi i h \frac{j}{k}} \\ &= \sum_{k=1}^{m} c(h,k),\end{aligned} \tag{116.5}$$

where

$$c(h,k) := \sum_{\substack{1\le j\le k\\ \gcd(j,k)=1}} e^{2\pi i j \frac{j}{k}} = \sum_{\substack{1\le j\le k\\ \gcd(j,k)=1}} \left(\zeta^j\right)^h \tag{w}$$

here $\zeta = e^{2\pi i/k}$. Note that

$$c(h,k) = \sum_{d|\gcd(h,k)} d\mu\left(\frac{k}{d}\right). \tag{116.7}$$

In particular,

$$c(1,k) = \mu(k). \tag{116.8}$$

For convenience we introduce a new function

$$M(x) = \sum_{n\le x} \mu(n). \tag{116.9}$$

We can use (116.7) to evaluate Σ_1. More precisely, we have

$$\begin{aligned}\Sigma_1 &= \sum_{k=1}^{m} c(h,k) \\ &= \sum_{k\le m} \sum_{d|\gcd(h,k)} d\mu\left(\frac{k}{d}\right) \\ &= \sum_{d|h} d \sum_{n\le \frac{m}{d}} \mu(n) \\ &= \sum_{d|h} d\mu\left(\frac{m}{d}\right).\end{aligned} \tag{116.10}$$

We can find a bound for Σ_1 as

$$|\Sigma_1| = \sum_{d|k} d\frac{m}{d} = m\sum_{d|k} 1 = md(h) = O(\sqrt{N}d(h)) \tag{116.11}$$

and the theorem follows from (116.11) and (116.4). □

Remark: The **Prime Number Theorem** is equivalent to $M(x) = O(x)$. More precisely, we know that

$$M(x) = O\left(\frac{x}{e^{\sqrt{\log(x)}}}\right) \tag{116.12}$$

which if employed in (116.10) would yield

$$\Sigma_1 = O\left(\frac{\sigma(h)\sqrt{N}}{he^{\sqrt{\log(N)}/k}}\right) \tag{116.13}$$

which would sharpen **Theorem 116** to

$$S_N = O\left(\frac{\sigma(h)\sqrt{N}}{he^{\sqrt{\log(N)}/k}} + \sqrt{N}\right) \tag{116.14}$$

which is uniform in h.

Remark: The Ramanujan sums $c(h,k)$ are multiplicative in h and k and are worth of independent study. See Apostal, pg. 160, for a discussion of $c(h,k)$.

15.8 Uniform Distribution of Sequences Using Weyl's Criterion

Remark: We have shown that $\{n\theta\}$ is uniformly distributed modulo 1 using Weyl's Criterion, as long as θ is irrational. I.M Vinogradov showed that $\{p\theta\}$ for θ irrational and p prime is uniformly distributed modulo 1 using his method of exponential sums combined with the Weyl Criterion. Using this method, Vinogradov showed that every sufficiently large odd integer is a sum of three primes.

Remark: Other interesting example is that if $\omega(n)$ is the number of prime divisors of n (either distinctly or with multiplicity, either way works) and if θ is irrational, then $\{\omega(n)\theta\}$ is uniformly distributed modulo 1. The Weyl sum for this problem is

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i h\omega(n)\theta}.$$

This of this as

$$\frac{1}{N}\sum_{n=1}^{N} z^{\omega(n)}$$

where

$$z := e^{2\pi i h\theta} \neq 1.$$

Note that $\omega(n)$ is an additive function, i.e., $\omega(mn) = \omega(m) + \omega(n)$. Thus, $z^{\omega(n)}$ is multiplicative. In the 1950s, Adle Selberg evaluated such sums.

Theorem 117: (van der Corput) Let $\{f(n)\}_{n\in\mathbb{N}}$ be a sequence of real number with the properties that the first differences $\Delta f(n) := f(n+1) - f(n)$ are monotone and that satisfies

$$\lim_{n\to\infty} \Delta f(n) = 0, \tag{117.1}$$

and

$$\lim_{n\to\infty} n|\Delta f(n)| = \infty. \tag{117.2}$$

Then, $\{f(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Proof: Begin by considering the following for $u, v \in \mathbb{R}$, using integration by parts:

$$\begin{aligned}\int_0^{u-v} (u-v-w)e^{2\pi iw}\, dw &= \left.\frac{e^{2\pi iw}(u-v-w)}{2\pi i}\right|_0^{u-v} + \frac{1}{2\pi i}\int_0^{u-v} e^{2\pi iw}\, dw \\ &= \frac{-(u-v)}{2\pi i} - \frac{e^{2\pi i}-1}{4\pi^2}\end{aligned} \tag{117.3}$$

Rewrite (117.3) as

$$4\pi^2\int_0^{u-v} (u-v-w)e^{2\pi iw}\, dw = 2\pi i(u-v) - e^{2\pi i(u-v)} + 1. \tag{117.4}$$

Thus,

$$\begin{aligned}4\pi^2\left|\int_0^{u-v} (u-v-w)e^{2\pi iw}\, dw\right| &= \left|2\pi i(u-v) - e^{2\pi i(u-v)} + 1\right| \\ &= \left|e^{2\pi iu} - e^{2\pi iv} - 2\pi i(u-v)e^{2\pi iv}\right|.\end{aligned} \tag{117.5}$$

On the other hand,

$$\begin{aligned}4\pi^2\left|\int_0^{u-v} (u-v-w)e^{2\pi iw}\, dw\right| &\leq 4\pi^2\left|\int_0^{u-v} (u-v-w)\, dw\right| \\ &= 4\pi^2\left|\int_0^{u-v} w\, dw\right| \\ &= 2\pi^2(u-v)^2.\end{aligned} \tag{117.6}$$

Thus, (117.5) and (117.6) yield

$$\left|e^{2\pi iu} - e^{2\pi iv} - 2\pi i(u-v)e^{2\pi iv}\right| \leq 2\pi^2(u-v)^2. \tag{117.7}$$

Given $h \in \mathbb{Z}$ with $h \neq 0$, put

$$u := hf(n+1) \text{ and } v := hf(n). \tag{117.8}$$

So, by (117.6) - (117.8), we have that

$$\left|\frac{e^{2\pi hf(n+1)}}{\Delta f(n)} - \frac{e^{2\pi ihf(n)}}{\Delta f(n)} - 2\pi ihe^{2\pi ihf(n)}\right| \leq 2\pi^2h^2\Delta f(n). \tag{117.9}$$

Thus,

$$\left|\frac{e^{2\pi hf(n+1)}}{\Delta f(n+1)} - \frac{e^{2\pi ihf(n)}}{\Delta f(n)} - 2\pi ihe^{2\pi ihf(n)}\right| \leq 2\pi^2h^2\Delta f(n) + \left|\frac{e^{2\pi ihf(n+1)}}{\Delta f(n+1)} - \frac{e^{2\pi if(n+1)}}{\Delta f(n)}\right|. \tag{117.10}$$

Now consider the Weyl sum

$$2\pi ih\sum_{n=1}^{N-1} e^{2\pi ihf(n)} = \sum_{n=1}^{N-1}\left(2\pi ihe^{2\pi ihf(n)} - \frac{e^{2\pi ihf(n+1)}}{\Delta f(n+1)} + \frac{e^{2\pi ihf(n)}}{\Delta f(n)}\right) + \left(\frac{e^{2\pi ihf(n+1)}}{\Delta f(n+1)} - \frac{e^{2\pi ihf(n)}}{\Delta f(n)}\right). \tag{117.11}$$

Hence,

$$\begin{aligned}\left|2\pi ih\sum_{n=1}^{N-1} e^{2\pi ihf(n)}\right| &\leq \sum_{n=1}^{N-1}\left|2\pi ihe^{2\pi ihf(n)} - \frac{e^{2\pi ihf(n+1)}}{\Delta f(n+1)} + \frac{e^{2\pi ihf(n)}}{\Delta f(n)}\right| + \left|\frac{e^{2\pi ihf(N)}}{\Delta f(N)} - \frac{e^{2\pi ihf(1)}}{\Delta f(1)}\right| \\ &\leq \sum_{n=1}^{N-1}\left|2\pi ihe^{2\pi ihf(n)} - \frac{e^{2\pi ihf(n+1)}}{\Delta f(n+1)} + \frac{e^{2\pi ihf(n)}}{\Delta f(n)}\right| + \frac{1}{\Delta f(N)} + \frac{1}{\Delta f(1)}. \end{aligned} \tag{117.12}$$

Thus from (117.9) - (117.12) we get that

$$\left|2\pi ih\sum_{n=1}^{N-1} e^{2\pi ihf(n)}\right| \leq 2\pi^2h^2\sum_{n=1}^{N-1}\Delta f(n) + \sum_{n=1}^{N-1}\left|\frac{1}{\Delta f(n+1)} - \frac{1}{\Delta f(n)}\right| + \frac{1}{|\Delta f(N)|} + \frac{1}{|\Delta f(1)|}. \tag{117.13}$$

Observe that since $\Delta f(n)$ is monotone:

$$\begin{aligned}\sum_{n=1}^{N-1}\left|\frac{1}{\Delta f(n+1)} - \frac{1}{\Delta f(n)}\right| &= \left|\sum_{n=1}^{N-1}\frac{1}{\Delta f(n+1)} - \frac{1}{\Delta f(n)}\right| \\ &= \left|\frac{1}{\Delta f(N)} - \frac{1}{\Delta f(1)}\right| \\ &\leq \left|\frac{1}{\Delta f(N)}\right| + \left|\frac{1}{\Delta f(1)}\right|. \end{aligned} \tag{117.14}$$

Thus (117.13) and (117.14) yield

$$\frac{1}{N-1}\left|\sum_{n=1}^{N-1} e^{2\pi ihf(n)}\right| \leq \frac{2\pi h}{N}\left|\sum_{n=1}^{N-1}\Delta f(n)\right| + \frac{1}{\pi h}\frac{1}{(N-1)^2} - \left(\frac{1}{|\Delta f(N)|} + \frac{1}{|\Delta f(1)|}\right). \tag{117.15}$$

Since $\Delta f(n) = O(1)$ as $n \to \infty$, the first term on the right-hand side of (117.15) is $O(1)$ (and goes to zero as $N \to \infty$). Since $n\Delta f(n) \to \infty$, the second term goes to zero as well. Thus, $f(n)$ satisfies **Weyl's Criterion**, Hence the fraction sequence $\{\{f(n)\}\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. $\square$

Corollary 117.1: (Fejer's Theorem) Let $f(x)$ be differentiable for $x \geq x_0$. If $f'(x) \to 9$ monotonically as $x \to \infty$ and if

$$\lim_{x\to\infty} x|f'(x)| = \infty. \tag{117.16}$$

then $\{f(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Corollary 117.2: Let $\alpha \in \mathbb{R}$ be nonzero and let $0 \leq \sigma < 1$, and let $\tau \in \mathbb{R}$. Then, $f(n) := \alpha n \left(\log(n)\right)^\tau$ is uniformly distributed modulo 1.

Theorem 118: The sequence $\{\log(n)\}_{n\in\mathbb{N}}$ is not uniformly distributed modulo 1.

Remark: $f(n) = \log(n)$ satisfies $\Delta f(n)$ is monotone and $\Delta f(n) \searrow 0$ as $n \to \infty$. But (117.2) is violated.

Proof: (version 1, using Weyl criterion) Let F be a smooth function. Consider

$$\begin{aligned}
\sum_{n=1}^{N} F(n) &= \int_{1^-}^{N^+} F(t)d[t] \\
&= \int_1^N F(t)dt - \int_{1^-}^{N^+} F(t)d\{t\} \\
&= \int_1^N F(t)dt - \left|F(t)\{t\}\right|_{1^-}^{N^+} + \int_1^N \{t\}F'(t)dt \\
&=
\end{aligned}$$

We take

$$F(t) := e^{2\pi i \log(t)}. \tag{118.2}$$

Now,

$$\left.\begin{aligned} F'(t) &= 2\pi i t^{2\pi i - 1} \\ \textstyle\int F(t)dt &= \dfrac{t^{2\pi i+1}}{2\pi i + 1} \end{aligned}\right\} \tag{118.3}$$

With this choice, we have

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i \log(n)} &= \frac{1}{N}\sum_{n=1}^{N} F(n) \\
&= \frac{1}{N}\int_1^N t^{2\pi i}dt + \frac{F(1)}{N} + \frac{2\pi i}{N}\int_1^N \{t\}t^{2\pi i - 1}dt \\
&= \left|\frac{t^{2\pi i+1}}{N(2\pi i+1)}\right|_1^N + \frac{F(1)}{N} + O\left(\frac{\log(N)}{N}\right) \\
&= \frac{N^{2\pi i}}{2\pi i+1} - \frac{1}{N(2\pi i+1)} + \frac{F(1)}{N} + O\left(\frac{\log(N)}{N}\right)
\end{aligned} \tag{118.4}$$

Note that the first term in (118.4) is not convergent as $n \to \infty$ but it is bounded, whereas the remaining three terms converge to zero. Thus,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i \log(n)} \neq 0$$

and hence it violates Weyl's criterion with $h = 1$. Therefore, $\{\log(n)\}_{n\in\mathbb{N}}$ is not uniformly distributed modulo 1. □

Remark: The above proof works for $f(n) = c\log(n)$ where $c \in \mathbb{R} \setminus \{0\}$. Taking $c = h \in \mathbb{Z} \setminus \{0\}$, we see that the Weyl criterion is violated for every h.

Proof: (version 2, by direct computation) Given $0 \leq a < b < 1$, when is

$$a \leq \{\log(n)\} < b? \tag{118.5}$$

This happens precisely when

$$e^{k+a} \leq n < e^{k+b} \tag{118.6}$$

where $k \in \mathbb{Z}$. Now, choose $N \in \mathbb{Z}$ large, and then we can find an m such that

$$e^{m+b} \leq N < e^{m+1}. \tag{118.7}$$

Then, the number of members of the sequence $\omega := \{\log(n)\}_{n\in\mathbb{N}}$ whose fractional part lies in $[a, b)$ is

$$A([a,b), N, \omega) = \sum_{k=1}^{m}((e^{k+b} - e^{k+a}) + O(1)) = (e^b - e^a)\frac{e^{m+1} - 1}{e - 1} + O(m). \tag{118.8}$$

Thus,

$$\frac{A([a,b), N, \omega)}{N} = \frac{e^b - e^a}{e - 1}\left(\frac{e^{m+1} - 1}{N}\right) + O\left(\frac{m}{N}\right). \tag{118.9}$$

Clearly, $m = O(\log(N))$ and so the second term on the right in (118.9) goes to zero, whereas the first term on the right in (118.9) foes not tend to any limit at $n \to \infty$. Hence, $\{\log(n)\}_{n\in\mathbb{N}}$ is not uniformly distributed modulo 1. □

Theorem 119: Suppose $\{f(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. Then,

$$\limsup_{n\to\infty} n|\Delta f(n)| = \infty. \tag{119.1}$$

We will establish this using a fundamental theorem.

Theorem T: (see Hardy's Divergent series, AMS Chelsea (1991), p. 121, Theorem 63) Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence such that

$$\sum a_n = A \quad (C, 1) \tag{119.2}$$

(i.e., it is Cesaro summable). Suppose that

$$a_n = O\left(\frac{1}{n}\right). \tag{119.3}$$

Then, $\sum a_n$ is convergent (to A, of course).

Remark: Let $A_n = a_1 + \cdots + a_n$. We say that $\sum a_n$ is (C,1) summable to A if

$$\frac{A_1 + A_2 + \cdots + A_n}{n} \to A$$

as $n \to \infty$.

Proof: Define

$$g(n) = e^{2\pi i f(n)} \tag{119.4}$$

and

$$a_n = \Delta g(n) = e^{2\pi i f(n+1)} - e^{2\pi i f(n)}. \tag{119.5}$$

The differences of the exponential function satisfy

$$|e^{2\pi iu} - e^{2\pi iv}| = \left|2\pi i \int_u^v e^{2\pi it} dt\right| \leq 2\pi |u - v|. \tag{119.6}$$

Thus,

$$a_n = \Delta g(n) = O(|f(n+1) - f(n)|) = O(\Delta f(n)). \tag{119.7}$$

Suppose that

$$\limsup_{n\to\infty} n|\Delta f(n)| < \infty. \tag{119.8}$$

Then,

$$|\Delta f(n)| = O\left(\frac{1}{n}\right). \tag{119.9}$$

Therefore,

$$a_n = O\left(\frac{1}{n}\right). \tag{119.10}$$

We now apply **Theorem T** with this choice of a_n. So the sequence of partial sums is

$$\begin{aligned} A_n &= a_1 + a_2 + \cdots + a_n \\ &= \sum_{m=1}^{n} \Delta g(m) \\ &= g(n+1) - g(1) \\ &= e^{2\pi i f(n+1)} - g(1). \end{aligned} \tag{119.11}$$

Hence

$$\lim_{N\to\infty} \frac{A_1 + \cdots + A_n}{N} = \left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i f(n+1)}\right) - g(1). \tag{119.12}$$

Since $\{f_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1, the first term on the right is zero. Thus,

$$\sum a_n = -g(1) \quad (C, 1).$$

Thus by **Theorem T**,

$$\sum a_n = -g(1)$$

converges. This clearly is false as seen by (119.11). Hence the theorem follows. □

15.9 A Theorem of van der Corput

Lemma 120: (van der Corput's Fundamental Inequality) Let $u_1, u_2, \ldots, u_N$ be complex numbers and $1 \leq H \leq N$, where H is arbitrary but fixed). Then,

$$H^2 \left| \sum_{n=1}^{N} u_n \right|^2 \leq H(N+H-1) \left[\sum_{n=1}^{N} |u_n|^2 \right] + 2(N+H-1) \left[\sum_{h=1}^{H-1} (H-h) \operatorname{Re} \left(\sum_{n=1}^{N-h} u_n \overline{u_{n+h}} \right) \right]. \quad (120.1)$$

Proof: If we set $u_n = 0$ for $n \leq 0$ and $n > N$, then we may write

$$H \sum_{n=1}^{N} u_n = \sum_{m=1}^{N+H-1} \sum_{h=0}^{H-1} u_{m-h}. \quad (120.3)$$

By applying the **Cauchy-Schwartz** inequality to (120.3), we get

$$\begin{aligned} H^2 \left| \sum_{n=1}^{N} u_n \right|^2 &\leq \left(\sum_{m=1}^{N+H-1} 1^2 \right) \left(\sum_{m=1}^{N+H-1} \left| \sum_{h=1}^{H-1} u_{m-h} \right|^2 \right) \\ &= (N+H-1) \sum_{m=1}^{N+H-1} \left(\sum_{h=1}^{H-1} u_{m-h} \right) \left(\sum_{h=1}^{H-1} \overline{u_{m-h}} \right) \\ &= (N+H-1) \left[\sum_{m=1}^{N+H-1} \sum_{h=1}^{H-1} |u_{m-h}|^2 \right] + 2(N+H-1) \operatorname{Re} \left(\sum_{m=1}^{N+H-1} \sum_{\substack{r,s=0 \\ s<r}}^{H-1} u_{m-r} \overline{u_{m-s}} \right) \\ &= (N+H-1) \left(\Sigma_1 + 2 \operatorname{Re}(\Sigma_2) \right). \end{aligned} \quad (120.4)$$

By an argument similar to (120.3), we have

$$\Sigma_1 = H \sum_{n=1}^{N} |u_n|^2. \quad (120.5)$$

With regard to Σ_2, we are summing terms of the form

$$u_n \overline{u_{n+h}} \quad (120.6)$$

with $h = r - s > 0$. But, the only contribution comes from values

$$1 \leq n \leq N - h.$$

With a little rearranging, we see that

$$\Sigma_2 = \sum_{h=1}^{H-1} \left[(H-h) \sum_{n=1}^{N-h} u_n \overline{u_{n+h}} \right]. \quad (120.7)$$

The Lemma follows form the values in (120.5) and (120.7). □

Theorem 120: (van der Corput's Difference Theorem) Let $\omega = \{x_n\}_{n \in \mathbb{N}}$ be a sequence of reals such that for each $h \in \mathbb{N}$, the sequence $\{x_{n_h} - x_n\}_{n \in \mathbb{N}}$ is uniformly distributed modulo 1. Then, $\{x_n\}_{n \in \mathbb{N}}$ is uniformly distributed modulo 1.

Proof: Let m be an arbitrary but fixed nonzero integer. Apply **Lemma 120** with

$$u_n = e^{2\pi i m x_n}, \quad |u_n| = 1 \quad (120.8)$$

and divide both sides of (120.1) by H^2N^2 to get

$$\left|\frac{1}{N}\sum_{m=1}^{N} e^{2\pi imx_n}\right|^2 \leq \frac{N+H-1}{HN} + 2\sum_{h=1}^{H-1}\left[\frac{(N+H-1)(H-h)(N-h)}{H^2N^2}\cdot\left|\frac{1}{N-h}\sum_{n=1}^{N-h} e^{2\pi im(x_n-x_{n+h})}\right|\right] \tag{120.9}$$

because $\mathrm{Re}(z) \leq |z|$ for all $z \in \mathbb{C}$.

Since $\{x_{n+h} - x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for each h, if we let $N \to \infty$ in (120.9) but keep H fixed, we get that

$$\limsup_{N\to\infty}\frac{1}{N}\left|\sum_{n=1}^{N} e^{2\pi imx_n}\right| \leq \frac{1}{\sqrt{H}}. \tag{120.10}$$

But, H is arbitrary! Hence, letting $H \to \infty$ in (120.10), we get that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} e^{2\pi imx_n} = 0 \tag{120.11}$$

for $m \in \mathbb{Z}\setminus\{0\}$. Thus by the **Weyl Criterion** the sequence $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. □

Theorem 121: (Hermann Weyl)

$$p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_m x^m$$

be a polynomial such that at least one of the coefficients α_j for $j > 0$ is irrational. Then, $\{p(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Remark: Herman Weyl proved this in his classic paper of 1916, but his proof was quite complicated. We will provide a simpler proof due to van der Corput.

Proof: (van der Corput) The case $m = 1$ is

$$p(x) = \alpha_1 x + \alpha_0$$

where α_1 irrational. Therefore,

$$\{p(n)\}_{n\in\mathbb{N}} = \{\alpha_1 n + \alpha_0\}_{n\in\mathbb{N}}. \tag{121.1}$$

We know that $\{\alpha_1 n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1, and hence so is $\{p(n)\}_{n\in\mathbb{N}}$ for all $\alpha_0 \in \mathbb{R}$.

We now prove the theorem by induction on $m := \deg(p)$. Let $m \geq 2$. We first consider the simpler case, namely

$$\alpha_1 \text{ irrational, and } \alpha_2, \alpha_3, \ldots, \alpha_m \in \mathbb{Q}. \tag{121.2}$$

(We don't really care about α_0.) In this case, write

$$p(x) = P(x) + \alpha_1 x + \alpha_0. \tag{121.3}$$

where $P(x)$ has only rational coefficients. Let L be defined to be the least common multiple of the denominators of $\alpha_2, \ldots, \alpha_m$. It then follows that

$$\{P(kL + d)\} = \{P(d)\}, \tag{121.4}$$

for $k \geq 0$ and any d. So, let $h \in \mathbb{Z} \setminus \{0\}$ and consider the Weyl sum for $p(m)$:

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i h p(m)} &= \frac{1}{N}\sum_{k=0}^{[N/L]-1}\sum_{d=1}^{L} e^{2\pi i h\{P(kL+d)+\alpha_1(kL+d)+\alpha_0\}} + \frac{1}{N}\sum_{n=([N/L]-1)\cdot L}^{N} e^{2\pi i h p(n)}. \\ &= \frac{1}{N}\sum_{d=1}^{L}\sum_{k=0}^{[N/L]-1} e^{2\pi i h\{p(kL+d)+\alpha_1(kL+d)+\alpha_0\}} + O\left(\frac{L}{N}\right) \\ &= \left(\sum_{d=1}^{L} e^{2\pi i h\{P(d)+\alpha_1 d+\alpha_0\}}\right)\left(\frac{1}{N}\sum_{k=0}^{[N/L]-1} e^{2\pi i h\alpha_1 Lk}\right) + O\left(\frac{L}{N}\right), \qquad (121.5)\end{aligned}$$

where the last equality is by virtue of (121.4).

Observe that in (121.5) the second term on the right is (ever better than) a Weyl sum for the sequence $\{\alpha_1 Lk\}_{k\in\mathbb{N}}$ with integer $h \in \mathbb{Z} \setminus \{0\}$. Therefore, the second factor in (121.5) tends to 0 as $N \to \infty$. The first fact is a finite sum, and therefore the product also goes to 0 as $N \to \infty$. Hence, $\{p(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 in this case, by the **Weyl Criterion**.

Now we consider the case where α_j is also irrational for some $j \geq 2$. Define q to be the largest integer for which α_q is irrational. We know the truth of the theorem for $q = 1$, and so we now argue by induction on q. For a fixed $h \in \mathbb{N}^+$ consider

$$p(x+h) - p(x) = \sum \beta_j x^j.$$

So, the largest j for which β_j is irrational is a value $j < q$. So, by the induction hypothesis, the sequence $\{p(n+h) - p(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for $n = 1, 2, \ldots$. But, $h \in \mathbb{Z}^+$ is arbitrary. So, by the **van der Corput Difference Theorem**, the sequence $\{p(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. □

Corollary: If $p(x) \in \mathbb{Z}[x]$ with degree ≥ 1, then $\{p(n)\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for each irrational θ.

Remark: The above methods can be combined to form more general theorems, one of which we prove now.

Theorem 122: Let $\omega := \{x_n\}_{n\in\mathbb{N}}$ be a sequence of reals such that

$$\Delta x_n := x_{n+1} - x_n \to \theta \qquad (122.1)$$

as $n \to \infty$, where θ is irrational. Then, $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Remark: van der Corput's **Theorem 117** does not apply to $\{n\theta\}_{n=1}^{\infty}$ with θ irrational, because that result required $\Delta x_n \to 0$, whereas here $\Delta x_n = \theta$.

Remark: Theorem 122 generalizes Weyl's original theorem on the uniform distribution modulo 1 of $\{n\theta\}_{n\in\mathbb{N}}$.

Proof: From the definition of Δx_n we have

$$x_n - x_m - (n-m)\theta = \sum_{j=m}^{n-1}(\Delta x_j - \theta). \qquad (122.2)$$

In view of (122.1), given $q \in \mathbb{Z}^+$ large, there exists $m_0 = m_0(q)$ such that

$$|\Delta x_j - \theta| < \frac{1}{q^2}, \quad j \geq m_0(q). \qquad (122.3)$$

Then,

$$|x_n - x_m - (n-m)\theta| \leq \sum_{j=m}^{n-1} |\Delta x_j - \theta| \leq \frac{n-m}{q^2}, \quad \text{for } n \geq j \geq m_0(q). \tag{122.4}$$

For an arbitrary but fixes nonzero integer h, we therefore have

$$|e^{2\pi i h x_n} - e^{2\pi i h(x_m + (n-m)\theta)}| \leq \frac{2\pi h(n-m)}{q^2}, \quad \text{for } n \geq m \geq m_0(q). \tag{122.5}$$

Thus,

$$\left| \sum_{n=m}^{m+q-1} e^{2\pi i h x_n} \right| \leq \left| \sum_{n=m}^{m+j-1} e^{2\pi i h(x_n + (n-m)\theta)} \right| + \sum_{n=m}^{m+q-1} \frac{2\pi h(n-m)}{q^2}. \tag{122.6}$$

Note that

$$\begin{aligned} \left| \sum_{n=m}^{m+q-1} e^{2\pi i h(x_n + (n-m)\theta)} \right| &= \left| \sum_{n=m}^{m+q-1} e^{2\pi i h n \theta} \right| \\ &= \frac{\left| e^{2\pi i q\theta} - 1 \right|}{|e^{2\pi i h\theta} - 1|} \\ &\leq \frac{2}{|e^{2\pi i h\theta} - 1|}. \end{aligned} \tag{122.7}$$

Now, from (122.5), (122.6), and (122.7), we get that

$$\left| \sum_{n=m}^{m+q-1} e^{2\pi i h x_n} \right| \leq \frac{2}{|e^{2\pi i h\theta} - 1|} + \pi h. \tag{122.8}$$

By breaking the range of summation into blocks of length q, we get

$$\left| \sum_{n=m}^{N} e^{2\pi i h x_n} \right| \leq \frac{N-M}{q} \left(\frac{2}{|e^{2\pi i h\theta} - 1|} + \pi h \right) + q. \tag{122.9}$$

Hence,

$$\left| \sum_{n=1}^{N} e^{2\pi i h x_n} \right| \leq m - 1 + O\left(\frac{N-m}{q} \cdot g \right) + q \tag{122.9}$$

yielding

$$\left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \right| \leq \frac{m-1}{N} + O\left(\frac{h}{q} \right) + \frac{q}{N}. \tag{122.10}$$

In (122.10) let $N \to \infty$ (keeping h and q arbitrary but fixed). Thus

$$\lim_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \right| = O\left(\frac{h}{q} \right). \tag{122.11}$$

Now since q can be chosen arbitrarily large, (122.11) implies that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} = 0. \tag{122.12}$$

This holds for every h. Hence by the **Weyl Criteron**, the sequence $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. □

15.10 Successive Differences

Define

$$\Delta x_n := x_{n+1} - x_n$$

and

$$\Delta^2 x_n = \Delta(\Delta x_n) = \Delta x_{n+1} - \Delta x_n = x_{n+2} - 2x_{n+1} + x_n.$$

More generally, with $\Delta^k x_n := \Delta(\Delta^{k-1} x_n)$ we have by induction that

$$\Delta^k x_n = \sum_{j=1}^{k} (-1)^j \binom{k}{k} x_{n+k-j}.$$

Theorem 123: (van der Corput) Let $\omega := \{f(n)\}_{n\in\mathbb{N}}$ be a sequence of reals such that for some $k \in \mathbb{Z}^+$, $\Delta^k f(n)$ is monotone,

$$\Delta^k f(n) \to 0 \text{ as } n \to 0, \tag{123.1}$$

and

$$\lim_{n\to\infty} n|\Delta^k f(n)| = \infty. \tag{123.2}$$

Then, the sequence $\{f(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Proof: We establish this by induction on k. We have already proved this for $k = 1$ in **Theorem 117**. So, assume the result is true for a certain k. Consider $\omega := \{f(n)\}_{n\in\mathbb{N}}$ satisfying the hypothesis for $k + 1$ for an arbitrary but fixed positive integer h. Consider

$$f(n + h) - f(n) = \sum_{j=0}^{h-1} \Delta f(n + j). \tag{123.3}$$

So,

$$\Delta^k(f(n + h) - f(n)) = \sum_{j=0}^{h-1} \Delta^{k+1} f(n + j). \tag{123.4}$$

Since $\Delta^{k+1} f$ is monotone, so is $\Delta^k f(n + j)$. Now since (123.1) holds for $k + 1$, we have from (123.4) that

$$\Delta^k(f(n + h) - f(n)) \to 0 \text{ as } n \to \infty. \tag{123.5}$$

Note also that

$$\left|\Delta^k(f(n + h) - f(n))\right| = \sum_{j=0}^{h-1} \left|\Delta^{k+1} f(n + h)\right| \tag{123.6}$$

due to the monotone property. Multiplying both sides by n, we get that the left-hand side goes to infinity by (123.2) for $k + 1$. Thus, by the induction hypothesis, $\{f(n + h) - f(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. But recall that h is arbitrary, and so $\{f(n)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 as well. □

Corollary: The sequence $\{n^\sigma\}_{n\in\mathbb{N}}$ for any nonintegral $\sigma > 0$ is uniformly distributed modulo 1.

15.11 Metric Theorems on Uniform Distribution

Theorem 124: (Koksma's Metric Theorem) Let $\omega := \{a_n\}_{n\in\mathbb{N}}$ be a sequence of distinct integers. Then, for almost all real θ, the sequence $\{a_n\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Remark: We are not assuming that $\{a_n\}_{n\in\mathbb{N}}$ is monotone (eventually), only that the a_n are distinct, i.e.,

$$|a_n - a_m \geq 1|$$

if $m \neq n$.

Proof: Clearly it suffices to show that $\{a_n\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $\theta \in [0,1]$, because the $\theta \in [N, N+1]$ for $N \in \mathbb{Z}$ are just translates of the $\theta \in [0,1]$ by N.

For nonzero $h \in \mathbb{Z}$ arbitrary but fixed, denote

$$S_h(N,\theta) := \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i h a_n \theta} \tag{124.1}$$

for $0 \leq \theta \leq 1$. Then

$$|S_h(N,\theta)|^2 = \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N} e^{2\pi i h(a_m - a_n)\theta} = S_h(N,\theta)S_h(N\theta). \tag{124.2}$$

Hence,

$$\begin{aligned}\int_0^1 |S_h(N\theta)|^2 d\theta &= \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\int_0^1 e^{2\pi i h(a_m-a_n)\theta}d\theta \\ &= \frac{1}{N^2}\cdot N \\ &= \frac{1}{N},\end{aligned} \tag{124.3}$$

since the integrals are nonzero only when $m = n$ and and in this case each such integral has value 1.

In particular,

$$\int_0^1 |S_h(N^2,\theta)|^2 d\theta = \frac{1}{N^2}. \tag{124.4}$$

This means that

$$\sum_{n=1}^{\infty}\int_0^1 |S_h(N^2,\theta)|^2 d\theta = \sum_{N=1}^{\infty}\frac{1}{N^2} < \infty. \tag{124.5}$$

Since the series is absolutely convergent, we may reverse the integration and summation by **Fatou's Lemma**to get that

$$\int_0^1 \left[\sum_{N=1}^{\infty}\left|S_h(N^2,\theta)\right|^2\right] d\theta < \infty. \tag{124.6}$$

The integral

$$\int_0^1 \sum_{N=1}^{\infty}\left|S_h(N^2,\theta)\right| d\theta$$

is defined as a Lebesgue integral, and so the convergence implies that

$$\sum_{N=1}^{\infty}\left|S_h(N^2,\theta)\right|^2 < \infty, \text{ almost everywhere.} \tag{124.7}$$

Thus,

$$\lim_{N\to\infty} \left|S_h(N^2,\theta)\right|^2 = 0 = \lim_{N\to\infty} S_h(N^2,\theta) \tag{124.8}$$

for almost all $\theta \in [0,1]$. Now, given an integer M arbitrarily large, determine N such that

$$N^2 \leq M < (N+1)^2. \tag{124.9}$$

Then,

$$\begin{aligned} S_h(M,\theta) &= \frac{1}{M}\sum_{n=1}^{M} e^{2\pi i h a_n \theta} \\ &= \frac{N^2}{M}\cdot\frac{1}{N^2}\sum_{n=1}^{N^2} e^{2\pi i h a_n \theta} + \frac{1}{M}\sum_{n=N^2+1}^{M} e^{2\pi i h a_n \theta}. \end{aligned} \tag{124.10}$$

Therefore,

$$|S_h(M,\theta)| \leq \frac{N^2}{M}\left|S_h(N^2,\theta)\right| + \frac{N}{M}. \tag{124.11}$$

Since $N = O\left(\sqrt{M}\right)$ and $N^2 \sim M$, the expression in (124.11) tends to zero as $M \to \infty$ almost everywhere in $[0,1]$. Since the countable intersection of almost everywhere sets in $[0,1]$ is an almost everywhere set in $[0,1]$, by applying this for $h = \pm 1, \pm 2, \cdots$, we reach the theorem. $\square$

Corollary 125: If $p(x) \in \mathbb{Z}[x]$ with $\deg(p) \geq 1$, then $\{p(n)\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $\theta \in \mathbb{R}$.

Remark: We know by **van der Corput's Theorem** that the set of θ is the set of irrationals.

Chapter 16

Diophantine Approximations and Transcendence

16.1 Koksma's General Metric Theorem

Remark: If $\omega := \{a_n\}_{n\in\mathbb{N}}$ is given, define u_ω to be the set of all θ for which $\{a_n\theta\}$ is uniformly distributed modulo 1. Then, **Koksma's Metric Theorem** says that

$$\mu(\mathbb{R} \setminus u_\omega) = 0.$$

Remark: If $\omega := \{b^m\}_{n\in\mathbb{N}}$ for $b \in \mathbb{Z}$ and $b > 1$, then (it will be shown later) u_ω is the set of reals whose expansion to base b is normal to base b. So, by **Koksma's Metric Theorem**, almost all numbers are normal to base b. Thus, almost all numbers are normal to every base (we call these numbers simply "normal"). Despite the preponderance of normal numbers, we have yet to produce a single one.

We extrapolate the principle ideas in the proof of **Koksma's Metric Theorem** to get the following generalization.

Theorem 126: Let $\{u_n(\theta)\}_{n\in\mathbb{N}}$ be an infinite sequence (of functional values depending on a parameter θ, Let

$$S_h(N,\theta) = \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i h u_n(\theta)} \tag{126.1}$$

and the integrals

$$I_h(N,a,b) = I_h(N) = \int_a^n |S_h(N,\theta)|^2 \, d\theta \tag{126.2}$$

where the $u_n(\theta)$ are Lebesgue measurable functions on $[a,b]$. If the series

$$\sum_{N=1}^{\infty} \frac{I_h(N)}{N} < \infty \tag{126.3}$$

for each nonzero $h \in \mathbb{Z}$, then the sequence $\{u_n(\theta)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $\theta \in [a,b]$.

Proof: Let nonzero $h \in \mathbb{Z}$ be arbitrary but fixed. In view of (126.3), we can find a function $\lambda(N) = \lambda_h(N)$ such that

$$\lambda(N) \text{ is increasing}, \quad \lambda(N) \to \infty, \tag{126.4}$$

and yet

$$\sum_{N=1}^{\infty} \frac{I_h(N)\lambda(N)}{N} < \infty. \tag{126.5}$$

(Proving this uses some ϵ-δ arguments. It is non-obvious, but not too tricky.)

Using the $\lambda(N)$, define a strictly increasing sequence of integers M_r by

$$M_{r+1} = \left[\frac{\lambda(M_r)}{\lambda(M_1) - 1} \cdot M_r \right] + 1. \tag{126.6}$$

We note immediately that since $\lambda(N) \to \infty$, we have

$$\frac{M_{r+1}}{M_r} \to 1 \text{ as } r \to \infty. \tag{126.7}$$

These M_r will play the role of N^2 in the proof of **Koksma's Metric Theorem**.

So, we decompose $[M, \infty)$ as

$$(M, \infty) = \bigcup_{r=1}^{\infty} (M_r, M_{r+1}]. \tag{126.8}$$

Consider an integer $N_r \in (M_r, M_{r+1}]$, so that

$$M_r < N_r \leq M_{r+1} \tag{126.9}$$

for which $I(N)$ attains its least value. We then have that

$$\begin{aligned} I(N_r) &\leq \frac{1}{M_{r+1} - M_r} \sum_{N=M_r+1}^{M_{r+1}} I(N) \\ &= \frac{1}{M_{r+1} - M_r} \sum_{N=M_r+1}^{M_{r+1}} N \cdot \frac{I(N)}{N} \\ &\leq \frac{M_{r+1}}{M_{r+1} - M_r} \sum_{N=M_r+1}^{M_{r+1}} \frac{I(N)}{N}. \end{aligned} \tag{126.10}$$

From the definition of M_{r+1} in (126.6) it follows that

$$\frac{M_{r+1} - M_r}{M_{r+1}} < \lambda(M_r). \tag{126.11}$$

Hence (126.10) and (126.11) yield

$$I(N_r) \leq \lambda(M_r) \sum_{N=M_r+1}^{M_{r+1}} \frac{I(N)}{N} \leq \sum_{N=M_r+1}^{M_{r+1}} \frac{I(N)\lambda(N)}{N} \tag{126.12}$$

because λ is increasing. By (126.12) and (126.5) we deduce that

$$\sum_{r=1}^{\infty} I(N_r) M\infty. \tag{126.13}$$

When combined with (126.2), this yields

$$\int_a^b \left(\sum_{r=1}^{\infty} |S_h(N_r, \theta)|^2 \right) d\theta < \infty. \tag{126.14}$$

Thus,

$$\sum_{r=1}^{\infty} |S_h(N_r, \theta)|^2 < \infty \tag{126.15}$$

almost everywhere in $[a, b]$. In particular,

$$\lim_{r\to\infty} S_h(N_r, \theta) = 0 \tag{126.16}$$

almost everywhere in $[a, b]$. This is true for each $h \neq 0$ and hence for all $h \in \mathbb{Z}^+$. In view of (126.7) and (126.9) it follows that

$$\lim_{r\to\infty} \frac{N_{r+1}}{N_r} = 1.$$

Now given a large integer N, determine r such that

$$N_r \leq N < N_{r+1} \tag{126.17}$$

and note that

$$|S_h(N, \theta)| \leq |S_h(N_r, \theta)| + \frac{N_{r+1} - N_r}{N_r} \to 0 \tag{126.18}$$

as $r \to \infty$. Thus, $\{u_n\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 almost everywhere in $[a, b]$. □

Remark: Koksma's Metric Theorem assumes that the members of the integer sequence $\{a_n\}_{n\in\mathbb{N}}$ are all distinct. **Theorem 126** allows us to handle the case where repetition occurs infinitely often. So, let $S(N)$ denote the number of pairs (m, n) with $1 \leq m, n \leq N$ for which $a_m = a_n$. Then,

$$I_h(N) = I(N) = \int_0^a |S_h(N, \theta)|^2 d\theta = \frac{S(N)}{N^2}. \tag{127.1}$$

Thus

$$\sum_{N=1}^{\infty} \frac{I(N)}{N} = \sum_{N=1}^{\infty} \frac{S(N)}{N^3}. \tag{127.2}$$

This leads to the following theorem.

Theorem 127: Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of integers such that there are $S(N)$ pairs (m, n) with $1 \leq m, n \leq N$ satisfying $a_n = a_m$. If the series

$$\sum_{N=1}^{\infty} \frac{S(N)}{N^3} < \infty$$

then $\{a_n\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all θ.

Remark: Theorems 126 & 127, which generalized **Koksma's Metric Theorem**, can be generalized further to deal with a wider class of sequences. For example, we will prove:

Theorem 128: (Koksma's General Metric Theorem) Let $\{u_n(x)\}_{n\in\mathbb{N}}$ be a sequence of real numbers defined for all $x \in [a, b]$. Let $u_n(x)$ be continuously differentiable on $[a, b]$. Suppose for $m \neq n$ that $u'_m(x) = u'_n(x)$ is monotone in x and that

$$|u'_m(x) - u'_n(x)| \geq k > 0 \tag{128.1}$$

where k is a constant which is independent of m, n, x. Then, $\{u_n(x)\}$ is uniformly distributed modulo 1 for almost all $x \in [a, b]$.

Remark: To prove **Theorem 128**, we will need to prove the following lemma:

Lemma 128: Let f be a real-valued function such that f' is monotone on $[a,b]$. Also, let

$$f'(x) \geq \lambda > 0 \qquad \text{or} \qquad f'(x) \leq -\lambda < 0 \tag{128.2}$$

for a constant λ. Let

$$J := \int_a^b e^{2\pi i f(x)}\, dx. \tag{128.3}$$

Then,

$$|J| < \frac{1}{\lambda}. \tag{128.4}$$

Proof of Lemma: First rewrite J as

$$J = \frac{1}{2\pi i}\int_a^b \frac{f'(x)2\pi i e^{2\pi i f(x)}}{f'(x)}dx = \frac{1}{2\pi i}\int_a^b \frac{1}{f'(x)}de^{2\pi i f(x)}. \tag{128.5}$$

Thus by the Mean Value Theorem, there exists $x_0 \in (a,b)$ such that

$$J = \frac{1}{2\pi i}\left(\frac{1}{f'(a)}\int_a^{x_0} de^{2\pi i f(x)} + \frac{1}{f'(b)}\int_{x_0}^b de^{2\pi i f(x)}\right). \tag{128.6}$$

Hence,

$$\begin{aligned} |J| &= \left|\frac{1}{2\pi i}\left(\frac{1}{f'(a)}\int_a^{x_0} de^{2\pi i f(x)} + \frac{1}{f'(b)}\int_{x_0}^b de^{2\pi i f(x)}\right)\right| \\ &\leq \frac{1}{2\pi}\left(\frac{2}{|f'(a)|} + \frac{2}{|f'(b)|}\right) \\ &\leq \frac{2}{\pi\lambda} \\ &< \frac{1}{\lambda}. \end{aligned} \tag{128.7}$$

This proves the Lemma. □

Proof of Theorem: For an arbitrary but fixed nonzero integer k, consider that

$$S_h(N,x) = \frac{1}{N}\sum_{n=1}^N e^{2\pi i h u_N(x)}, \tag{128.8}$$

for $x \in [a,b]$. Then,

$$\begin{aligned} I_h(N) &= \int_a^b |S_h(N,x)|^2 dx \\ &= \frac{1}{N^2}\sum_{m=1}^N\sum_{n=1}^N \int_a^b e^{2\pi i h(u_m(x)-u_n(x)}dx \\ &\leq \frac{1}{N^2}\sum_{m=1}^N\sum_{n=1}^N \left|\int_a^b e^{2\pi i h(u_m(x)-u_n(x))}dx\right| \\ &\leq \frac{b-a}{N} + \frac{2}{N^2}\sum_{m=2}^N\sum_{n=1}^{m-1}\left|\int_a^b e^{2\pi i h(u_m(x)-u_n(x))}dx\right|, \end{aligned} \tag{128.9}$$

where the first term on the right-hand side of (128.9) is from the diagonal terms $m = n$ and the last term is from $m \neq n$. By **Lemma 128** and (128.1), we have that

$$\left|\int_a^b e^{2\pi i h(u_m(x)-u_n(x))}dx\right| \leq \frac{1}{|h|}\max\left(\frac{1}{|u_m'(a)-u_n'(a)|}, \frac{1}{|u_m'(b)-u_n'(b)|}\right) \tag{128.10}$$

by the monotone property. Thus by (128.9) and (128.10), we get that

$$I_h(N) \leq \frac{b-a}{N} + \frac{2}{|h|N^2} \sum_{m=2}^{N} \sum_{n=1}^{m-1} \left(\frac{1}{|u'_m(a) - u'_n(a)|} + \frac{1}{|u'_m(b) - u'_n(b)|} \right), \tag{128.11}$$

where in (128.11) we have used

$$\max(\alpha, \beta) \leq |\alpha| + |\beta|.$$

From (128.1) we see from the inner term that

$$|u'_m(a) - u'_n(a)| \geq |m-n| \cdot k, \tag{128.12}$$

owing to the monotonicity, and because

$$\begin{aligned} |u'_m(a) - u'_n(a)| &= |(u'_m(a) - u'_{m-1}(a) + u'_{m-1}(a) - u'_{m-2}(a) + \cdots + u'_{n+1}(a) - u'_n(a)| \\ &= |u'_m(a) - u'_{m-1}(a)| + |u'_{m-1}(a) - u'_{m-2}(a)| + \cdots + |u'_{n+1}(a) - u'_n(a)| \\ &\geq \underbrace{k + k + \cdots + k}_{m-n \text{ times}} = |m-n| \cdot k. \end{aligned} \tag{128.13}$$

We have a similar lower bound for $|u'_m(b) - u'_n(b)|$. Thus (128.11) and (128.12) yield

$$\begin{aligned} I_h(N) &\leq \frac{b-a}{N} + \frac{2}{N^2} \sum_{m=2}^{N} \sum_{j=1}^{m-1} \frac{1}{jk} \\ &\leq \frac{b-a}{N} + \frac{2}{kN^2} \sum_{m=2}^{N} \log(m) \\ &= O\left(\frac{b-a}{N} + \frac{\log(N)}{kN} \right). \end{aligned} \tag{128.14}$$

This upper bound for $I_h(N)$ implies

$$\sum_{N=1}^{\infty} \frac{I_h(N)}{N} < \infty \tag{128.15}$$

and so **Theorem 128** follows from **Theorem 126**. □

Remark: Theorem 128 has several striking corollaries.

Corollary 128-1: Let $\{\gamma_n\}_{n\in\mathbb{N}}$ be a sequence of reals such that for some $\delta > 0$,

$$|\lambda_m - \lambda_n| \geq \delta > 0 \text{ for } m \neq n.$$

Then, $\{\gamma_n x\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $x \in \mathbb{R}$.

Proof: Apply **Theorem 128** to $u_n(x) := \lambda_n x$ with $x \in [k, k+1]$, with $k \in \mathbb{Z}$ arbitrary but fixed. This shows that $\{\lambda_n x\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $x \in [r, r+1]$. Thus, $\{\lambda_n x\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $x \in \mathbb{R}$. □

Corollary 128-2: The sequence $\{x^n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $x > 1$.

Proof: Set $u_n(x) := x^n$ is uniformly distributed modulo 1 for almost all $x > 1$. Thus $u'_n(x) = nx^{n-1}$ and this is clearly monotone in the differences

$$u'_{n+1}(x) = u'_n(x) = (n+1)x^n - nx^{n-1} = nx^{n-1}(x-1) + x^n. \tag{128.16}$$

Also,

$$|u'_{n+1}(x) - u'_n(x)| > x > 0.$$

Thus by **Theorem 128**, $\{u_n(x)\}_{n\in\mathbb{N}} = \{x^n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $x > 1$. □

Remark: Corollary 128-2 can be extended easily as follows.

Corollary 128-3: Let $\delta > 0$ be constant and $F(n)$ for $n = 1, 2, \ldots$ be a sequence of values > 1 satisfying $|F(m) - F(n)| \geq \delta$ for $m \neq n$. Let $\lambda \neq 0$ be an arbitrary constant. Then, $\{\lambda x^{F(n)}\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $x > 1$.

Proof: We concentrate on the interval $[k, k+1]$ with $k \in \mathbb{Z}^+$ and $k \geq 1$ arbitrary but fixed. Set

$$u_n(x) = \lambda x^{F(m)}.$$

Then,

$$u'_m(x) - u'_n(x) = \lambda \left(F(m)x^{F(m)-1} - F(n)x^{F(n)-1} \right) \tag{128.17}$$

and

$$u''_m(x) - u''_n(x) = \lambda \left(F(m)(F(m)-1)x^{F(m)-2} - F(n)(F(n)-1)x^{F(n)-2} \right). \tag{128.18}$$

Note that (128.18) implies

$$\operatorname{sgn}\left(u''_m(x) - u''_n(x)\right) = \operatorname{sgn}\left(\lambda(F(m) - F(n))\right) \tag{128.19}$$

and so $u'_m(x) - u'_n(x)$ is monotone for $x \in [k, k+1]$. Also, (128.17) yields

$$|u'_m(x) - u'_n(x)| = |\lambda| x^{-1} x^{\min\{F(m),F(n)\}} \left(F(m)x^{F(m)-\min\{F(m),F(n)\}} - F(n)x^{F(n)-\min\{F(m),F(n)\}} \right). \tag{128.10}$$

Clearly one of $F(m) - \min\{F(m), F(n)\}$ or $F(n) - \min\{F(m), F(n)\}$ is 0. Say that that first one is 0. Then, the expression in (128.20) is

$$|u'_m(x) - u'_n(x)| = |\lambda| x^{F(m)-1} \left| F(m) - F(n)x^{F(n)-F(m)} \right|. \tag{128.21}$$

Now, $x \geq 1$ since $x \in [k, k+1]$ and so we have that

$$\left| F(m) - F(n)x^{F(n)-F(m)} \right| = F(n)x^{F(n)-F(m)} - F(m)$$

(in this case) attains its minimum value at $x = k$, and this values is $\geq$ the value at $x = 1$. Thus,

$$\left| F(m) - F(n)x^{F(n)-F(m)} \right| \geq |F(m) - F(n)|, \quad x \in [k, k+1] \tag{128.22}$$

which combined with (128.21) and the hypothesis $|F(m) - F(n)| \geq \delta$ for $m \neq n$ yields,

$$|u'_m(x) - u'_n(x)| \geq |\lambda|\delta, \quad x \in [k, k+1]. \tag{128.23}$$

So, by **Theorem 128**, we see that $\{u_n(x)\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 for almost all $x \in [k, k+1]$ and hence for almost all $x > 1$, by writing $[1, \infty)$ as a countable union of $[k, k+1]$ for $k = 1, 2, 3, \ldots$. □

16.2 The Pisot-Vijayaraghavan Numbers

Definition: An algebraic integer α is called a Pisot-Vijayaraghavan number, or PV for short, if $\alpha > 1$ but all of its conjugates have absolute value < 1.

Example: The golden mean $\frac{1+\sqrt{5}}{2}$ has conjugate $\frac{1-\sqrt{5}}{2}$, and so is PV.

Example: The silver mean $1+\sqrt{2}$ has conjugate $1-\sqrt{2}$ and so is PV.

Theorem 129: If θ is a PV number, then $\|\theta^n\| \to 0$ as $n \to \infty$. I.e., $\{\theta^n\}$ clusters around 0.

Proof: Let $\theta_2, \theta_3, \ldots, \theta_d$ be the conjugates of $\theta =: \theta_1$. Since θ is an algebraic integers, we have for each $n \in \mathbb{Z}^+$,

$$\theta_1^n + \theta_2^n + \cdots + \theta_d^n =: f(n) \in \mathbb{Z}. \tag{129.1}$$

Since $|\theta_j| < 1$, we must have that

$$|\theta_j^n \to 0 \tag{129.2}$$

as $n \to \infty$. Hence,

$$\theta_2^n + \cdots + \theta_d^n \to 0 \tag{129.3}$$

as $n \to \infty$. Hence we will have that

$$\|\theta^n\| \to 0 \tag{a}$$

s $n \to \infty$. □

Remark: Both 0 and 1 can occur as cluster points for a given θ. Indeed with

$$\theta := \alpha = \frac{1+\sqrt{5}}{2}$$

we have

$$-1 < \theta' = \frac{1-\sqrt{5}}{2} < 0$$

and so

$$\theta'^n \to 0$$

but alternates in sign. Thus 0 and 1 are both cluster points of $\{\alpha^n\}$. The same is true for $1-\sqrt{2}$.

History and Nomenclature: The first person to discover these numbers was **Axel Thue** in 1912 , and he was aware of **Theorem 129**. **G. H. Hardy** discussed these numbers in 1919 quite independently, but from the point of view of Diophantine Approximations. (J. Indian Math. Soc., 1919). Then, **Charles Pisot** proved a partial converse:

Theorem 129-P: If α is an algebraic number greater than 1 for which there is $\lambda \in \mathbb{R}$ such that

$$\|\lambda\theta^n\| \to 0 \tag{129.3}$$

as $n \to \infty$. Then, θ is a PV number.

Remark: Pisot "extended" this to all complex numbers, i.e., he dropped the assumption that α was algebraic, subject to a slightly stronger convergence condition.

Theorem 129-P*: If $\theta \in \mathbb{R}$ and $\theta > 1$ for which there is a $0 \neq \lambda \in \mathbb{R}$ such that

$$\sum_{n=1}^{\infty} \|\lambda\theta^n\|^2 < \infty \tag{129.4}$$

then θ is a PV number.

Remark: In view of **Theorem 129-P***, it is conjectured that in **Theorem 129-P** the assumption that θ is algebraic can be dropped. It is still an open problem whether a transcendental $\tau > 1$ exists for which $\|\tau^n\| \to 0$ as $n \to \infty$.

16.3 Normal Numbers

Motivation: Normal numbers were first studied by Emil Borel. Given $\omega = \langle b^n \rangle$, where $b \geq 2$ is an integer. We know that

$$u_\omega = \{\theta \in \mathbb{R} \mid \langle b^n \theta \rangle \text{ is uniformly distributed modulo } 1\}$$

is almost all of $\mathbb{R}$ (from Koksma).

Question: Determine u_ω. Given an integer base b, the b-adic expansion of any real α to the base b is

$$\alpha = [\alpha] + \{\alpha\} = [\alpha] + \sum_{n=1}^{\infty} \frac{a_n}{b^n} \tag{130.1}$$

where

$$a_n \in \mathbb{Z}, \quad 0 \leq a_n \leq b-1 \tag{130.2}$$

and

$$a_n < b-1 \text{ infinitely often.} \tag{130.3}$$

Remark: Condition (130.3) is to confirm uniqueness of the decimal expansion and also to be consistent with $\{\alpha\} \in [0, 1)$.

Definition: A real number α is <u>simply normal</u> if

$$\lim_{N \to \infty} \frac{\nu_b(a; N)}{N} = \frac{1}{b} \tag{130.4}$$

where

$$\nu_b(a; N) := \text{ the number of } a_i \text{ for } i \leq N \text{ satisfying } a_i = a. \tag{130.5}$$

If different α are to be considered simultaneously, we need to write $\nu_b(a; N; \alpha)$.

Definition: More generally, consider any block B_k of k digits $b_1 b_2 \cdots b_k$ and

$$\nu_b(B_k; N; \alpha) := \text{ the number of occurrences of } B_k \text{ among the first } N \text{ digits of } \alpha. \tag{130.6}$$

We say that α is <u>normal to base b</u> if

$$\lim_{N \to \infty} \frac{\nu_b(B_k; N; \alpha)}{N} = \frac{1}{b^k} \tag{130.7}$$

for every block of k digits.

Example: An example of a number which is simply normal to base 10 but not normal is

$$.\overline{0123456789}$$

This is clearly simply normal, but it's not noaml because, for example, the block 22 never occurs.

Theorem: If α is normal to base b, then α is irrational.

Remark: Simply normal numbers can be both rational and irrational.

Theorem 131: α is normal to base b if and only if $\langle b^n\alpha\rangle$ is uniformly distributed modulo 1.

Proof: Let the b-adic expansion of α be given by (130.1). So, the series has value $\{\alpha\}$. Consider any block B_k of k digits

$$B_k : b_1 \cdots b_k. \tag{131.1}$$

This block coincides with a block of k digits of α, namely

$$b_1 \cdots b_k = a_m a_{m+1} \cdots a_{m+k-1} \tag{131.2}$$

if and only if

$$\alpha = [\alpha] + \left[\sum_{n=1}^{m-1} \frac{a_n}{b^n}\right] + \frac{b_1}{b^m} + \frac{b_2}{b^{m+1}} + \cdots + \frac{b_k}{b^{m+k-1}} + \left[\sum_{n=m+k}^{\infty} \frac{a_n}{b^n}\right]. \tag{131.3}$$

Multiplying (131.3) by b^{m-1} we obtain

$$\begin{aligned}\{b^{m-1}\alpha\} &= \frac{b_1}{b} + \frac{b_2}{b^2} + \cdots + \frac{b_k}{b^k} + \sum_{n=m+k}^{\infty} \frac{a_n}{b^{n-m+1}} \\ &= \frac{b_1 b^{k-1} + b_2 b^{k-2} + \cdots + b_k}{b^k} + \sum_{n'=k+1}^{\infty} \frac{a_{n'+m-1}}{b^{n'}},\end{aligned} \tag{131.4}$$

where $n' = n - m + 1$. Thus,

$$0 \le \sum_{n=k+1}^{\infty} \frac{a_{n'+m-1}}{b^{n'}} < \sum_{n'=k+1}^{\infty} \frac{b-1}{b^{n'}} = \frac{b-1}{b^{k+1}}(1 - \frac{1}{\delta}) = \frac{1}{b^k}. \tag{131.5}$$

Thus

$$\{b^{m-1}\alpha\} \in \left[\frac{b_1 b^{k-1} + \cdots + b_k}{b^k}, \frac{b_1 b^{k-1} + \cdots + b_k + 1}{b^k}\right] \qquad = I_k. \tag{131.6}$$

Thus, (131.6) implies that

$$A(I(B_k); N - k + 1; \omega) = \nu_b(B_k; N; \alpha). \tag{131.7}$$

If $\{\{b^n\alpha\}\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1, then

$$\lim_{N\to\infty} \frac{\nu_b(B_k; N; \alpha)}{N} = \lim_{N\to\infty} \frac{A(I(B_k); N - k + 1; \omega)}{N} = |I(B_k)| = \frac{1}{b^k}, \tag{131.8}$$

by (131.6). Since this holds for every B_k, we conclude that α is normal to base b.

Conversely if α is normal to base b, then (131.8) holds in the following manner:

$$\lim_{N\to\infty} \frac{A(I(B_k); N - k + 1; \omega)}{N} = \lim_{N\to\infty} \frac{\nu_b(B_k; N + k - 1; \alpha)}{N} = \frac{1}{b^k}. \tag{131.9}$$

This means that

$$\lim_{N\to\infty} \frac{A([a, b); N; \omega)}{N} = b - a$$

holds for $[a, b)$ of type $I(B_k)$. So, it holds for all finite unions of such $I(B_k)$ and these will yield all intervals $[a, b)$ with rational endpoints. Every subinterval $[a, b) \subseteq [0, 1)$ can be approximated by intervals with rational endpoitns. Hence (131.9) holds for all subintervals. Thus, $\{b^n\alpha\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. □

Corollary: (Borel) Almost all numbers are normal to base b.

Corollary: (Borel) Almost all numbers are normal to every base. Such numbers are called normal, or absolutely normal.

Remark: An example of a normal number to base 10 is

$$\alpha = .01234567891011121314\ldots. \tag{131.10}$$

If we construct such a number to base 9, we would get

$$\alpha' = .012345678101121314\ldots, \tag{131.11}$$

but $\alpha \neq \alpha'$ and α' is normal to base 9.

Remark: It is an open problem to construct a number which is normal to all bases. It is conjectured that e, π, and $\sqrt{2}$ are such numbers.

Remark: Normal numbers were first studies by E. Borel in 1914 before Weyl's paper on uniform distibution modul 1. **Theorem 131** is due to Wall in his PhD thesis.

16.4 Uniform Distribution of Integer Sequences

Remark: This subject was studied comparitively recently, first by I. Niven.

Definition: An integer sequence $\{a_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m (for $m \in \mathbb{Z}^+$ with $m \geq 2$) if

$$\lim_{N\to\infty} \frac{A(j;m;N)}{N} = \frac{1}{m} \tag{132.1}$$

where $j = 1, 2, \ldots, m-1$ and

$$A(j;m;N) = \sum_{\substack{1\leq i\leq N \\ a_i\equiv j \pmod m}} 1. \tag{132.2}$$

We say that $\{a_n\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$ if $\{a_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m for all $m \in \mathbb{Z}^+$ with $m \geq 2$.

Remark: We begin with a Weyl-type criterion for uniform distribution in $\mathbb{Z}$.

Theorem 132: A necessary and sufficient condition that $\{a_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m is

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h a_n/m} = 0 \tag{132.3}$$

for $h = 1, 2, \ldots, m-1$.

Proof: We begin by noting that

$$\frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h a_n/m} = \frac{1}{N} \sum_{j=0}^{m-1} \sum_{\substack{1\leq n\leq N \\ a_n\equiv j \pmod m}} e^{2\pi i h a_n/m}. \tag{132.4}$$

If $\{a_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m, then by (132.4) we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h a_n/m} = \sum_{j=0}^{m-1} e^{2\pi i h j/m} \left[\lim_{N\to 0} \frac{A(j;m;N)}{N}\right] = \frac{1}{m} \sum_{j=0}^{m-1} \left(e^{2\pi i h/m}\right)^j = \frac{\zeta^m - 1}{\zeta - 1} = 0, \tag{132.5}$$

where $\zeta := e^{2\pi i h/m} \neq 1$, and so the necessity of (132.3) follows.

To prove sufficiency, we use inversion ideas, namely

$$\sum_{h=0}^{m-1} e^{2\pi i h(a_i - j)/m} = \begin{cases} 1, & a_i \equiv j \pmod m \\ 0, & a_i \not\equiv j \pmod m \end{cases} \tag{132.6}$$

So, (132.6) yields for a fixed j:

$$\begin{aligned}\frac{A(j;m;N)}{N} &= \frac{1}{N} \sum_{\substack{1\le n\le N\\ a_i\equiv j \pmod m}} 1\\ &= \frac{1}{N}\sum_{1\le i\le N}\frac{1}{m}\sum_{h=0}^{m-1} e^{2\pi i h(a_i-j)/m}\\ &= \frac{1}{m}\sum_{h=0}^{m-1}\frac{e^{-2\pi i hj/m}}{N}\sum_{1\le i\le N} e^{2\pi i h a_i/m}\\ &= \frac{1}{m}+\frac{1}{m}\sum_{h=1}^{m-1} e^{-2\pi i hj/m}\frac{1}{N}\sum_{1\le i\le N} e^{2\pi i h a_i/m}. \qquad (132.7)\end{aligned}$$

As $N\to\infty$, the inner sum on the right tends to 0 by view of (132.3). Thus as $N\to\infty$, (132.1) holds. Thus, $\{a_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m. □

16.5 Connection Between Uniform Distribution Mod 1 of Reals and Uniform Distribution Mod m of Integer Sequences

Theorem 133: Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of reals such that for some fixed $m \in \mathbb{Z}$ with $m \geq 2$ the sequence $\omega := \{x_n/m\}_{n\in\mathbb{N}}$ is uniformly distribued modulo 1. Then, $\{[x_n]\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m.

Proof: For $m \geq 2$ and any integer j satisfying $0 \leq j \leq m-1$, we have

$$\begin{aligned}
[x_n] \equiv j \pmod{m} &\iff [x_n] = mk + j, \text{ for some } k \in \mathbb{Z} \\
&\iff x_n = mk + j + \alpha, \text{ with } \alpha \in [0,1) \\
&\iff \frac{x_n}{m} = k + \frac{j}{m} + \frac{\alpha}{m}, \text{ with } \alpha \in [0,1) \\
&\iff \left\{\frac{x_n}{m}\right\} \in \left[\frac{j}{m}, \frac{j+1}{m}\right). \qquad (133.1)
\end{aligned}$$

Thus,

$$A(j;m;N) = A\left(\left[\frac{j}{m}, \frac{j+1}{m}\right); n; \omega\right). \qquad (133.2)$$

Hence

$$\lim_{N\to\infty} \frac{A(j;m;N)}{N} = \lim_{N\to\infty} \frac{A\left(\left[\frac{j}{m}, \frac{j+1}{m}\right); N; \omega\right)}{N} = \frac{1}{m}. \qquad (133.3)$$

Therefore, $[x_n]$ is uniformly distributed modulo m. □

Theorem 133*: If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of reals such that for every integer $m \geq 2$ the sequence $\{\frac{x_n}{m}\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1, then $\{[x_n]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$.

Theorem 134: (Niven) Let θ be irrational. Then, $\{[n\theta]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$.

Proof: Let $x_n := n\theta$. Since θ is irrational, so is $\frac{\theta}{m}$ for all integers m. So by Weyl's theorem, the sequence $\{n\frac{\theta}{m}\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1, i.e., $\{\frac{x_n}{m}\}_{n\in\mathbb{N}}$ is uniformly distributed with m arbitrary. So, by **Theorem 133***, we conclude that $\{[n\theta]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$.

Corollary 134: Let θ be irrational and let $|\theta| < 1$. Define an increasing sequence of integers $\{a_n\}_{n\in\mathbb{N}}$ such that there is an integer between $a_n\theta$ and $(a_n+1)\theta$, for all n. Then, $\{a_n\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$.

Proof: Without loss of generality, we have that $0 < \theta < 1$. Consider the sequence, $\{j\theta \_ j \in \mathbb{N}$ of all integral multiples of θ. Since θ is irrational, the sequence has the following two properties:

- Every integer $n \geq 1$ occurs between two consecutive members of the sequence.

- Between two consecutive numbers there is at most one integer.

Then, for all $n \in \mathbb{N}$, we can determine $a_n \in \mathbb{Z}$ uniquely such that

$$a_n\theta < n < (a_n+1)\theta. \qquad (136.3)$$

The correspondence $n \leftrightarrow a_n$ is a bijection, rewrite (134.3) as

$$a_n < n\theta^{-1} < a_n + 1 \qquad (134.4)$$

and so

$$a_n = [n\theta^{-1}].$$

Thus, $\{a_n\} = \{[n\theta^{-1}]\}$ is uniformly distributed in $\mathbb{Z}$. □

Theorem 135: For $\theta \in \mathbb{R}$, $\{[n\theta]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$ if and only if θ is irrational or $\theta = \frac{1}{d}$ for $0 \neq d \in \mathbb{Z}$.

Proof: We only need to discuss the $\theta \in \mathbb{Q}$ case. So, let $\theta = \frac{a}{b}$ for $a, b \in \mathbb{Z}$ and $\gcd(a, b) = 1$ with $b \geq 1$. Write

$$a = bq + r, \text{ where } 0 \leq r < b. \tag{135.1}$$

Then,

$$n\theta = n\frac{a}{b} = nq + \frac{nr}{b}$$

and so

$$[n\theta] = nq + \left[\frac{nr}{b}\right]. \tag{135.2}$$

Replacing n by $n + b$ in (135.2), we get

$$\begin{aligned}[(n+b)\theta] &= nq + bq + \left[\frac{(n+b)r}{b}\right] \\ &= nq + bq + r + \left[\frac{nr}{b}\right] \\ &= a + nq + \left[\frac{nr}{b}\right] \\ &= a + [n\theta]. \end{aligned} \tag{135.3}$$

Thus, b is a period (modulo $|a|$) for this sequence. If we now assume that $\{[n\theta]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$, then this forces

$$|a| \mid b$$

But, $\gcd(a, b) = 1$ and therefore $a = 1$, i.e., $\theta = \frac{1}{d}$ for $d \in Z$.

Conversely, if $\theta = \frac{1}{d}$, then without loss of generality, $d \in \mathbb{Z}^+$, and so

$$\{[n\theta]\}_{n\in\mathbb{N}} = \underbrace{0, 0, \ldots, 0}_{d \text{ times}}, \underbrace{1, 1, \ldots, 1}_{d \text{ times}}, \underbrace{2, 2, \ldots, 2}_{d \text{ times}}, \ldots,$$

and this is clearly uniformly distributed in $\mathbb{Z}$. □

Remark: If $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed moduloe 1, then for all $m \in \mathbb{Z}^+$, $\{mx_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 because $\{mx_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if for each $0 \neq h \in \mathbb{Z}$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{\infty} e^{2\pi i h(mx_n)} = 0.$$

But then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{\infty} e^{2\pi i h(mx_n)} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{\infty} e^{2\pi i (hm)x_n} = 0$$

by applying the **Weyl Criterion** to $\{x_n\}_{n\in\mathbb{N}}$ with $h \mapsto hm$.

But, the converse is not true. For example, let $x_n := \{n\theta\}$ with θ irrational This is uniformly distributed modulo 1 and $x_n \in [0, 1)$. But,

$$\frac{x_n}{m} \in \left[0, \frac{1}{m}\right)$$

which is clearly not uniformly distributed modulo 1 for $m \geq 2$ because not all of $(0, 1)$ is covered.

Theorem 134-P: Let θ be irrational. Then, $\{[p\theta]\}_{p \text{ prime}}$ is uniformly distributed in $\mathbb{Z}$.

Proof: Set $x_p := p\theta$. Since $\frac{\theta}{m} \in \mathbb{R} \setminus \mathbb{Q}$, we have

$$\langle \frac{x_p}{m} \rangle = \langle p\frac{\theta}{m} \rangle,$$

which is uniformly distributed modulo 1. □

Question: Analogous to **Theorem 135**, we may ask whether $\{[p\theta]\}$ is uniformly distributed in $\mathbb{Z}$ for any rational θ. The answer is "no".

Example: Let $\theta = \frac{1}{2}$. Then, $[p\theta] = \frac{p-1}{2}$ for $p = 3, 5, \ldots$.

Question: If $\{a_k\}_{k\in\mathbb{N}}$ is unformly distirbuted modulo m and uniformly distributed modulo n, with $\gcd(m,n) = 1$, then is $\{a_k\}_{k\in\mathbb{N}}$ uniformly distributed modulo mn? Again, the answer is "no".

Example: Let $m = 2$ and $n = 3$. Define:

$$a_k := \begin{cases} k, & k \equiv 0, 1, 2, 5 \pmod 6 \\ k-2, & k \equiv 3 \pmod 6 \\ k+2, & k \equiv 4 \pmod 6 \end{cases} \tag{136.1}$$

Then,

$$\{a_k\}_{k\in\mathbb{N}} = (\overline{0, 1, 2, 1, 6, 5})$$

and this is uniformly distributed modulo 2 and 3, but not modulo 6.

Remark: This sequence actually has the stronger property that it is uniformly distributed modulo p^α for every power of a prime p, but is not uniformly distributed modulo $\mathbb{Z}$.

Theorem 136: A sequence $\omega := \{x_n\}_{n\in\mathbb{N}}$ of reals is uniformly distributed modulo 1 if and only if the sequence $\{[mx_n]\}_{n\in\mathbb{N}}$ of integers is uniformly distributed modulo m for each $m \geq 2$.

Proof: Define the sequence ω_m by

$$\omega_m := \{mx_n\}_{n\in\mathbb{N}}$$

for $m = 2, 3, \ldots$. So, for a given m, the statement $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 is the same as saying that

$$\frac{1}{m}\omega_m = \left\{ \frac{mx^n}{m} \right\}_{n\in\mathbb{N}}$$

is uniformly distributed modulo 1. Thus by **Theorem 133**, we get that $\{[mx_n]\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m. This holds for each $m \geq 2$.

Conversely, let $\{[mx_n]\}_{n\in\mathbb{N}}$ be uniformly distributed modulo m for each $m \geq 2$. For an arbitrary but fixed m, we know (see (133.2)) that

$$\lim_{N\to\infty} \frac{A\left(\left[\frac{j}{m}, \frac{j+1}{m}\right]; N; \omega\right)}{N} = \lim_{N\to\infty} \frac{A(j; m; N)}{N} = \frac{1}{m} = \left|\left[\frac{j}{m}, \frac{j+1}{m}\right)\right| \tag{136.1}$$

for $j = 0, 1, 2, \ldots, m-1$.

So, by taking finite unions over $m \geq 2$ and j, we see from (136.1) that

$$\lim_{N\to\infty} \frac{A([\alpha, \beta); N; \omega)}{N} = \beta - \alpha \tag{136.2}$$

for all subintervals of $[0, 1)$ with *rational* end points α and β. Since every subinterval of $[0, 1)$ can be approximated by intervals with rational end points, we see that (136.2) holds for *all* subintervals. Hence ω is uniformly distributed modulo 1. □

Theorem: 137 (A Quantitaive Version of **Theorem 132**) For any sequence of integers $\{a_n\}_{n\in\mathbb{N}}$, and any pairs of integers m, N, we have that

$$\sum_{h=1}^{m} -1 \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i j a_n/m} \right|^2 = m \sum_{j=0}^{m-1} \left(\frac{A(j;m;N)}{N} - \frac{1}{m} \right)^2. \tag{131.7}$$

Corollary: Theorem 132 is a corollary, letting $N \to \infty$ in (137.1).

Proof: Begin by expanding the left-hand side of (137.1) to get

$$\begin{aligned} \sum_{n=1}^{m-1} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h a_n/m} \right|^2 &= \frac{1}{N^2} \sum_{h=1}^{m-1} \left[\sum_{n=1}^{N} \sum_{\ell=1}^{N} e^{2\pi i h(a_n - a_\ell)/m} \right] \\ &= \frac{1}{N^2} \sum_{\substack{1\le n\le N \\ 1\le \ell\le N \\ a_n \not\equiv a_\ell \pmod m}} \left[\sum_{h=1}^{m-1} e^{2\pi i h(a_n-a_\ell)} \right] + \frac{1}{N^2} \sum_{\substack{1\le n\le N \\ 1\le \ell\le N \\ a_n \equiv a_\ell \pmod m}} \left[\sum_{h=1}^{m-1} e^{2\pi i h(a_n-a_\ell)} \right] = \frac{1}{N^2} (\Sigma_1 + \end{aligned} \tag{137.2}$$

respectively. Note that

$$\Sigma_2 = (m-1) \sum_{\substack{1\le n\le N \\ 1\le \ell\le N \\ a_n \equiv a_\ell \pmod m}} 1 \tag{137.3}$$

and that the *inner sum* in Σ_1 is

$$\sum_{h=1}^{m-1} e^{2\pi i h(a_n-a_\ell)/m} = \left[\sum_{h=0}^{m-1} e^{2\pi i h(a_n-a_\ell)/m} \right] - 1 = -1. \tag{137.4}$$

So, (137.2), (137.3), and (137.4) yield

$$\sum_{h=1}^{m-1} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h a_n/m} \right|^2 = \frac{1}{N^2} \left[- \sum_{\substack{1\le n\le N \\ 1\le \ell\le N \\ a_n \not\equiv a_\ell \pmod m}} 1 + (m-1) \sum_{\substack{1\le n\le N \\ 1\le \ell\le N \\ a_n \equiv a_\ell \pmod m}} \right] = \frac{1}{N^2} \left[- \sum_{\substack{1\le n\le N \\ 1\le \ell\le N}} 1 + m \sum_{\substack{1\le n\le N \\ 1\le \ell\le N \\ a_n \equiv a_\ell \pmod m}} 1. \right. \tag{137.5}$$

The second sum in (137.5) is

$$\sum_{\substack{1\le n\le N \\ 1\le \ell\le N \\ a_n \equiv a_\ell \pmod m}} 1 = \sum_{j=0}^{m-1} 1 \sum_{\substack{1\le n\le N \\ a_n \equiv j \pmod m}} 1 \sum_{\substack{1\le \ell\le N \\ a_\ell \equiv a_n \pmod N}} 1 = \sum_{j=0}^{m-1} A^2(j;m;N). \tag{137.6}$$

Therefore,

$$\sum_{j=0}^{m-1} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h a_n/m} \right|^2 = -1 + m \sum_{j=0}^{m-1} \frac{A^2(j;m;N)}{N^2}. \tag{137.7}$$

Expanding the right-hand side of (137.1), we get

$$\begin{aligned} m\sum_{j=0}^{m-1}\left(\frac{A(j;m;N)}{N}-\frac{1}{m}\right)^2 &= m\sum_{j=0}^{m-1}\left(\frac{A^2(j;m;N)}{N^2}-\frac{2A(j;m;N)}{Nm}+\frac{1}{m^2}\right) \\ &= m\sum_{j=0}^{m-1}\frac{A^2(j;m;N)}{N^2}-\frac{2}{N}\sum_{j=0}^{m-1}A(j;m;N)+1 \\ &= m\sum_{j=0}^{m-1}\frac{A^2(j;m:N)}{N^2}-1 \end{aligned} \tag{137.8}$$

which is the same as (137.7). □

Theorem 138: Let α be irrational and $\theta \in \mathbb{R}$. Then

(i) If $0 \neq \theta \in \mathbb{Q}$, then $\{[n\theta]\alpha\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

(ii) If θ is irrational, then $\{[n\theta]\alpha\}_{n\in\mathbb{N}}$ is uniformly distributred modulo 1 if and only if $1, \theta, \theta\alpha$ are linearly independent over $\mathbb{Q}$.

Proof of (i) Let $\theta := r/s$ for $r \neq 0$ and $r, s \in \mathbb{Z}$ with $s \geq 1$ and $\gcd(r, s) = 1$. Define

$$M := \left[\frac{N}{s}\right]$$

where N is a large positive integer. So, for $0 \neq h \in \mathbb{Z}$, we have

$$\begin{aligned} \sum_{n=1}^{N} e^{2\pi i h[n\theta]\alpha} &= \sum_{n=1}^{Ms} e^{2\pi i h[n\theta]\alpha} + O(1) \\ &= \sum_{k=0}^{M-1}\sum_{m=1}^{s} e^{2\pi i h\left[\frac{ks+m}{s}\cdot r\right]\alpha} + O(1) \\ &= \sum_{m=1}^{s} e^{2\pi i h\left[\frac{mr}{s}\right]\alpha}\sum_{k=0}^{M-1} e^{2\pi i h k r\alpha} + O(1). \end{aligned} \tag{138.1}$$

Note that $hr\alpha$ is irrational, and so the inner sum in (138.1) is

$$\sum_{k=0}^{M-1} e^{2\pi i h k r\alpha} = \frac{e^{2\pi i M h r\alpha}-1}{e^{2\pi i h r\alpha}-1} = O(1) \tag{138.2}$$

because the denominator is never equal to zero. Thus by **Weyl's Criterion**, we get that $\{[n\theta]\alpha\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. □

Proof of (ii): Let θ be irraional. Suppose that $1, \theta, \theta\alpha$ satisfy the linear dependence relation

$$u + v\theta = w\theta\alpha \tag{138.4}$$

for $u, v, w \in \mathbb{Z}$ not all zero. Note that in (138.4), we have $w \neq 0$ because $w = 0$ implies that $\theta \in \mathbb{Q}$, which is not true. With this non-zero $w \in \mathbb{Z}$, we will show that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i w[n\theta]\alpha} \neq 0 \tag{138.5}$$

i.e., the Weyl criterion fails for this choise of $h = w$. Hence, $\{[n\theta]\alpha\}_{n\in\mathbb{N}}$ will not be uniformly distributed modulo 1. From (138.4), we infer that

$$\begin{aligned} e^{2\pi i w[n\theta]\alpha} &= e^{2\pi i w(n\theta - \{n\theta\})\alpha} \\ &= e^{2\pi i(n(w\theta)\alpha - w\{n\theta\}\alpha)} \\ &= e^{2\pi i(n(u+v\theta) - w\{n\theta\}\alpha)} \\ &= e^{2\pi i(v-w\alpha)\{n\theta\}} \\ &= g(\{n\theta\}), \end{aligned} \tag{138.6}$$

where

$$g(x) := e^{2\pi i(v-w\alpha)\{x\}}. \tag{138.7}$$

Clearly, g is periodic of period 1 and therefore

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i w[n\theta]\alpha} &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g(n\theta) \\ &= \int_0^1 g(x)dx \\ &= \int_0^1 e^{2\pi i(v-w\alpha)x} dx \\ &= \frac{e^{2\pi i(v-w\alpha)-1}}{2\pi i(v-w\alpha)} \end{aligned} \tag{138.8}$$

provided that $v - w\alpha \neq 0$. Next, if $0 \neq v - w\alpha \in \mathbb{Z}$, then $\theta = -u/(v-w\alpha) \in \mathbb{Q}$, which is a contradiction. Hence the limit in (138.8) is nonzero. So, the Weyl criterion is isolated and so $[n\theta]\alpha$ is not uniformly distributed modulo 1.

FInally, suppose that $1, \theta, \theta\alpha$ are linearly independent over $\mathbb{Q}$. For $0 \neq h \in \mathbb{Z}$, write

$$e^{2\pi i h[n\theta]\alpha} = e^{2\pi i h n\theta - h\{n\theta\}\alpha} =: f(n\theta, n\theta\alpha) \tag{138.9}$$

where

$$f(x, y) := e^{2\pi i(hy - h\{x\}\alpha)}. \tag{138.10}$$

Analogous to **Kronecker's Theorem**, we have that the sequence $\{(n\theta, n\theta\alpha)\}$ is uniformly distributed modulo the unit square and therefore by a 2-dimensional version, we have

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h[n\theta]\alpha} &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(n\theta, n\theta\alpha) \\ &= \int_0^1 \int_0^1 f(x, y)dxdy \\ &= \int_0^1 \int_0^1 e^{2\pi i(hy - hx\alpha)} dxdy \\ &= \left[\int_0^1 e^{2\pi i h y} dy\right] \left[\int_0^1 e^{-2\pi i h x\alpha} dx\right] \\ &= 0 \cdot \left[\int_0^1 e^{-2\pi i h x\alpha} dx\right] = 0, \end{aligned} \tag{138.11}$$

because $h \neq 0$. Hence, $\{[n\theta]\alpha\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. $\square$

Lemma 139-1: If $\{x_n\}_{n\in\mathbb{N}}$ is a real sequence that is uniformly distributed modulo 1, and if $\{y_n\}_{n\in\mathbb{N}}$ is a sequence such that

$$\lim_{n\to\infty}(x_n - y_n) = \alpha, \tag{139.1}$$

then $\{y_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Proof: Exercise.

Remark: This lemma includes the situation

$$x_n - y_n = \epsilon_n \to 0 \text{ as } n \to \infty. \tag{139.2}$$

Lemma 139-2: Let $\{x_n\}_{n\in\mathbb{N}}$ be a real sequence that is uniformly distributed modulo 1, and let $\{[x_n]\}_{n\in\mathbb{N}}$ be uniformly distributed modulo m for some integer $m \geq 2$. Let $\{y_n\}$ be a real sequence such that (139.2) holds. Then, $\{[y_n]\}_{n\in\mathbb{N}}$ is uniformly distributed modulo m. Thus, if $\{[x_n]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$, so is $\{[y_n]\}_{n\in\mathbb{N}}$.

Remark: A condition like $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 is necessary because if no condition is assumed on $\{x_n\}_{n\in\mathbb{N}}$, we can construct a counterexample: let $x_n = n$ for $n \in \mathbb{N}$. Then, $x_n = [x_n]$, which is uniformly distributed in $\mathbb{Z}$. In particular, it is uniformly distributed modulo 2. Define

$$y_n := \begin{cases} n + \dfrac{1}{n+1}, & n > 0 \text{ odd} \\ n - \dfrac{1}{n+1}, & n > 0 \text{ even} \end{cases}.$$

Then,

$$[y_n] : 1, 1, 3, 3, 5, 5, \ldots.$$

16.6 An Application

Theorem: 139 Let F_n be the Fibonacci sequence:

$$F_0 = 0, \;\; F_1 = 1, \;\; F_n = F_{n-1} + F_{n-2}, \;\; n \geq 2.$$

Then,

(i) $\omega := \{\{\log(F_n)\}\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

(ii) $\sigma := \{[\log(F_n)]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$.

Proof: We use Binet's formula

$$F_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}, \tag{139.3}$$

with

$$\alpha := \frac{1+\sqrt{5}}{2}, \quad \beta := \frac{1-\sqrt{5}}{2}, \quad \alpha - \beta = \sqrt{5}.$$

We know that α is a PV number because $|\beta| < 1$. Hence,

$$\begin{aligned} \log(F_n) &= \log(\alpha^n - \beta^n) - \log(\sqrt{5}) \\ &= \log(\alpha^n) + \log\left(1 - \frac{\beta^n}{\alpha^n}\right) - \log(\sqrt{5}) \\ &= n\log(\alpha) + \log\left(1 - \frac{\beta^n}{\alpha^n}\right) - \log(\sqrt{5}). \end{aligned} \tag{139.4}$$

Since $\alpha \neq 0$ is algebraic, $\log(\alpha)$ is transcendental and hence irrational. Thus,

$$x_n := n\log(\alpha) \tag{139.5}$$

is uniformly distributed modulo 1. Hence $y_n := \log(F_n)$ satisfies (139.1). Hence $\log(F_n)$ is uniformly distributed modulo 1.

Next with x_n as above, we know that $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1, and $\{[x_n]\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{Z}$. Additionally, $y_n - x_n$ converges to some limit, and so by the lemma and (139.2),

$$[y_n] = [\log(F_n)]$$

is uniformly distributed in $\mathbb{Z}$. □

Appendix A

Frank Patane: Irrationality Measure

Recall: Let α be irrational, written in simple continued fraction form as $\alpha = [a_0; a_1, a_2, \cdots]$, and define $M(\alpha) = \limsup_{n\in\mathbb{N}} ([a_{n+1}; a_{n+2}, \ldots] + [0; a_n, a_{n-1}, \ldots, a_0])$.

Theorem: There are uncountably many real numbers α with $M(\alpha) = 3$.

Proof: We will write 1_j for a succession of j terms each 1. Similarly, we write 2_j for j successive terms of 2. Let $r_1, r_2, \cdots$ be a strictly increasing sequence of positive integers and consider

$$\alpha := [1; 1_{r_1}, 2, 2, 1_{r_2}, 2, 2, 1_{r_3}, 2, 2, \ldots].$$

Define

$$\beta_k := \alpha_{k+1} + \frac{q_{k-1}}{q_k} = [a_{k+1}; a_{k+2}, \ldots] + [0; a_k, \ldots, a_0].$$

If $a_{k+1} = 1$, then

$$\alpha_{k+1} < 2 \qquad \text{and} \qquad \frac{q_{k-1}}{q_k} < 1.$$

So, in this case, $\beta_k < 3$.

If k runs through indices such that $a_k = a_{k+1} = 2$, then

$$\beta_k = [2; 1, 1, \ldots] + [0; 2, 1, 1, \ldots].$$

So,

$$\lim_{k\to\infty} \beta_k = [2; \overline{1}] + [0; 2, \overline{1}] = \left(2 + \frac{1}{\left[\frac{1+\sqrt{5}}{2}\right]}\right) + \frac{1}{2 + \frac{1}{\left[\frac{1+\sqrt{5}}{2}\right]}} = 3.$$

In the last case, if $a_k = a_{k+1} = 2$, then we have

$$\beta_k = [2; 2, 1, 1, \ldots] + [0; 1, 1, \ldots],$$

which equals the same sum as above with the two terms switched.

Thus for any sequence of strictly increasing positive integers $\{r_i\}$, we get $M(\alpha) = 3$, where $\alpha = [1; 1_{r_1}, 2, 2, 1_{r_2}, 2, 2, 1_{r_3}, 2, 2, \ldots]$. So, we have a bijection between these sequences and irrational numbers α with $M(\alpha) = 3$.

Let α, α' be of the previous form for sequences r_i and r_i', respectively. Then, we have $\alpha \sim \alpha'$ if and only if r_i and r_i' have the same tails at some point (perhaps with a different starting point from each sequence).

Suppose toward a contradiction that the inequivalent sequences (among all strictly increasing sequences at positive integers) can be listed as $R_1, R_2, \ldots$, and take $R_1 = 1, 2, 3, \cdots$. If $i > 1$, then $R_i \not\sim R_1$, and so R_i misses infinitely many positive integers (otherwise R_1 and R_i would eventually agree).

For each R_i, for $i > 1$, we form the complement S_i, which is the strictly increasing sequence of integers not occuring in R_i. Define the sequence $T = \{T_1, T_2, \ldots\}$. Pick $T_1 \in S_2$, and pick $T_2 \in S_3$ so that $T_1 < T_2$. Pick

$$\begin{array}{llll} 1 + T_2 < T_3 \in S_2, & T_3 < T_4 \in S_3, & T_4 < T_5 \in S_4, & \\ 1 + T_5 < T_6 \in S_2, & T_6 < T_7 \in S_3, & T_7 < T_8 \in S_4, & T_8 < T_9 \in S_5, \\ 1 + T_9 < T_{10} \in S_2, & \cdots & & \end{array}$$

Clearly, T is a strictly increasing sequence of positive integers, and T has infinitely many T_i which are contained in S_k for all $k \geq 2$. Therefore, T has infinitely many terms T_i that are not in R_k, for all $k \geq 2$. So, $T \not\sim R_k$, for $k \geq 2$. By forcing $T_3 > 1 + T_2$, $T_6 > 1 + T_5$, etc, we ensure that $T \not\sim R_1$, since T misses infinitely many positive integers, because for each $T_{\frac{i(i+1)}{2}}$, this term is more than 1 above its predecessor. Thus T is not in our list, hence we have a contradiction, thus the number of irrationals with Markov constant 3 is uncountable. $\square$

If $\alpha \not\sim \dfrac{1+\sqrt{5}}{2}$ (i.e., the tail of α is not all 1s) and $\alpha \not\sim \sqrt{8}$ (i.e., the tail of α is not all 2s), then how small can $M(\alpha)$ be? Such an α must have infinitely many $\ldots, 2, 1, \ldots$ appearing in the expansion. If we have infinitely many $\ldots, 2, 1, 2, \ldots$, then we can make the appropriate cut so that $M(\alpha) \geq 3$. Similarly for the pattern $\ldots, 1, 2, 1, \ldots$.

In general, if we have $\ldots, 1, 2_{2k+1}, 1, \ldots$ or $\ldots, 2, 1_{2k+1}, 2, \ldots$, appearing, then $M(\alpha) \geq 3$.

In the remaining cases, we have $\ldots, 2, 2, 1, 1, \ldots$ infinitely often, so $M(\alpha) \geq [2; 1, 1, \ldots] + [0; 2, \ldots]$. Minimum of all these possibilities is $, \ldots, 1, 1, 2, 2, 1, 1, 2, 2, \ldots$, in which case $M(\alpha) = \dfrac{\sqrt{221}}{5}$, which the smallest possible.

Appendix B

Todd Molnar: Periodic Continued Fractions

Definition: A periodic continued fraction is a simple continued fraction $x = [a_0, a_1, a_2, \ldots]$ in which $a_\ell = a_{\ell+k}$ for a fixed positive $k \in \mathbb{Z}$ and all $\ell \geq L$. The set of partial quotients $a_L, a_{L+1}, \ldots, a_{L+k-1}$ is called the period. We will write this as $[a_0, a_1, \ldots, a_{L-1}, \overline{a_L, a_{L+1}, \ldots, a_{L+k-1}}]$.

Theorem 1: (Euler, 1737) A period continued fraction is a quadratic irrational, i.e., the irrational root of some quadratic equation with integers coefficients.

Proof: Let $x := [a_0, a_1, \ldots, a_{L-1}, \overline{a_L, a_{L+1}, \ldots, a_{L+k-1}}]$. Let

$$\begin{aligned} a'_L &:= [a_L, a_{L+1}, \cdots] \\ &= [a_L, a_{L+1}, \cdots, a_{L+k-1}, a_L, a_{L+1}, \cdots] \\ &= [a_L, a_{L+1}, \cdots, a_{L+k-1}, a'_L]. \end{aligned}$$

By a result from class,

$$a'_L = \frac{p' a'_L + p''}{q' a'_L + q''}$$

where $\frac{p'\,'}{q}$ and $\frac{p''}{q''}$ are the last two convergents of

$$[a_L, a_{L+1}, \ldots, a_{L+k-1}].$$

So,

$$q' + a'^2_L + (q'' - p')a'_L - p'' = 0.$$

But,

$$x = \frac{p_{L-1} a'_L + p_{L-2}}{q_{L-1} a'_L - q_{L-2}},$$

and hence

$$a'_L = \frac{p_{L-2} - q_{L-2} x}{q_{L-1} x - p_{L-1}},$$

and so

$$q' \left(\frac{p_{L-2} - q_{L-2} x}{q_{L-2} x - p_{L_1}} \right)^2 + (q'' - p') \left(\frac{p_{L-2} - q_{L-2} x}{q_{L-2} x - p_{L_1}} \right) - p'' = 0.$$

Thus, this is a solution to $ax^2 + bx + c$, with $a, b, c \in \mathbb{Z}$ and $x \notin \mathbb{Q}$. Also observe that $b^2 - 4ac \neq 0$.

Theorem 2: (Lagrange, 1770 - Converse to Theorem 1) A simple continued fractions which represents a quadratic irrational is periodic.

Proof: If $x = [a_0, a_1, \ldots, a'_n]$, with a'_n the "tail", then since x is a quadratic irrational, we have that $ax^2 + bx + c = 0$ for some $a, b, c \in \mathbb{Z}$, where $b^2 - 4ac \neq 0$. Furthermore,

$$x = \frac{p_{n-1}a'_n + p_{n-1}}{q_{n-1}a'_n + q_{n-2}}.$$

We see that $A_n a'^2_n + B_n a'_n + C_n = 0$, where

$$\begin{aligned} A_n &= ap^2_{n-1} + bp_{n-1}q_{n-1} + cq^2_{n-1} \\ B_n &= 2ap_{n-1}p_{n-2} + b(p_{n-1}q_{n-2} + p_{n-2}q_{n-1}) + 2cq_{n-1}q_{n-2} \\ C_n &= ap^2_{n-2} + bp_{n-2}q_{n-2} + cq^2_{n-2} \end{aligned}$$

Note that $C_n = A_{n-1}$ and $A_n, B_n, C_n \in \mathbb{Z}$.

So, if $A_n = 0$, then $ax^2 + bx + c = 0$ has root p_{n-1}/q_{n-1} which is a rational root and is not possible. Hence $A_n \neq 0$. Therefore, we actually do have a quadratic equation. Furthermore, $A_n y^2 + B_n y + C_n = 0$ is an equation with at least one root equal to a'_n. A straightforward but tedious calculation tells us that

$$B^2_n - 4A_nC_n = (b^2 - 4ac)(p_{n-1}q_{n-2} - p_{n-2}q_{n-1})^2 = b^2 - 4ac \neq 0.$$

(We used the fact that $p_{n-1}q_{n-2} - p_{n-2}q_{n-1} = 1$.)

Recall that since p_n and q_n are the convergents:

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q^2_n}.$$

Hence, there exists a real δ_{n-1} with $|\delta_{n-1}| < 1$, such that

$$p_{n-1} = xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}}.$$

Therefore,

$$\begin{aligned} A_n &= a\left(xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}}\right)^2 + bq\left(xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}}\right) + cq^2_{n-1} \\ &= (ax^2 + bx + c)q^2_{n-1} + 2ax\delta_{n-1} + a\frac{\delta^2_{n-1}}{q^2_{n-1}} + b\delta_{n-1} \\ &= 2ax\delta_{n-1} + a\frac{\delta^2_{n-1}}{q^2_{n-1}} + b\delta_{n-1}. \end{aligned}$$

Thus, $|A_n| < 2|ax| + |a| + |b|$. Similarly, $|C_n| < 2|ax| + |a| + |b|$.

Now, $B^2_n - 4A_nC_n < 4(2|ax| + |a| + |b|)^2 + |b^2 - 4ac|$, thus A_n, B_n, C_n are all bounded in absolute value by numbers independent of n¿ There are only a finite number of different triples (A_n, B_n, C_n) such that $A_n y^2 + B_n y + C_n = 0$. We may find a triplet (A, B, C) which occurs three times, say as

$$(A_{n_1}, B_{n_1}, C_{n_1}), \quad (A_{n_2}, B_{n_2}, C_{n_2}), \quad (A_{n_3}, B_{n_3}, C_{n_3}).$$

Substituting these values into an earlier equation gives that $a'_{n_1}, a'_{n_2}, a'_{n_3}$ are all roots of $Ay^2 + By + C = 0$. So, as numbers, two of those are the same, and by the uniqueness of expansion, we can say that two of them are equal as continued fractions. Without loss of generality, say $a'_{n_1} = a'_{n_2}$, and therefore

$$[a_{n_1}, a_{n_1+1}, \ldots] = [a_{n_2}, a_{n_2+1}, \ldots].$$

Thus $a_{n_2} = a_{n_1}$, and $a_{n_2+1} = a_{n_1+1}$, etc. Hence the continued fraction is periodic. □

Appendix C

Duc Huynh: Ford's Theorem

[Some of talk missing.]

Ford's Theorem: Given any irrational complex number α, there exists infinitely many rational complex numbers $\frac{u}{v}$ such that

$$\left|\alpha - \frac{u}{v}\right| < \frac{1}{\sqrt{3}|v|^2} = \frac{1/\sqrt{3}}{|v|^2}.$$

Furthermore, the constant $\sqrt{3}$ is the best possible in the sense that the result becomes false for any larger constant.

Proof will follow after some additional Theorems and Lemmas.

Theorem 4.1: Given any irrational complex number α, there exists infinitely many rational complex numbers $\frac{u}{v}$ such that

$$\left|\alpha - \frac{u}{v}\right| < \frac{2}{|v|^2}.$$

Proof: Let $n \in \mathbb{N}$ be a large positive integer. Let v run through the $(n+1)^2$ complex integers $a+bi$, with $0 \le a \le n$ and $0 \le b \le n$. For each v, we select $u \in \mathbb{Z}[i]$ such that $\alpha v - u = x + yi$, for $0 \le x < 1$ and $0 \le y < 1$. Assume toward a contradiction that the pairs are not distinct, i.e., $\alpha v_1 - u_1 = \alpha v_2 - u_2$, which implies that α is not an irrational complex number, a contradiction. Hence the pairs are distinct. Additionally, these pairs are all inside the unit square.

Dividing the unit square up into $\frac{1}{n} \times \frac{1}{n}$ squares, we see that only one pair can lie on a corner - if two pairs did, then we could find a rational complex α. By construction, $|(\alpha v_1 - u_1) - (\alpha v_2 - u_2)| < \frac{\sqrt{2}}{n}$ for some two pairs. Now, $|v_1 - v_2| \le n\sqrt{2}$, and this is equivalent to $n \ge \frac{|v_1 - v_2|}{\sqrt{2}}$. Hence we get that

$$\left|\alpha - \frac{u_1 - u_2}{v_1 - v_2}\right| < \frac{\sqrt{2}}{n|v_1 - v_2|} < \frac{2}{|v_1 - v_2|^2}.$$

Let $u = u_1 - u_2$ and $v = v_1 - v_2$. We have that

$$\left|\alpha - \frac{u}{v}\right| < \frac{\sqrt{2}}{|v|^2},$$

and

$$\left|\alpha - \frac{u}{v}\right| < \frac{\sqrt{2}}{n|v|} < \frac{2}{|v|^2}.$$

Suppose we have only finitely many $\frac{u}{v}$ such that $|\alpha - \frac{u}{v}| < \frac{\sqrt{2}}{v}$, we can choose $\frac{u_0}{v_0}$ such that $\left|\alpha - \frac{u_0}{v_0}\right| = \epsilon$ is smallest.

Claim: Let $a_1, \ldots, a_k \in \mathbb{R}^+$, such that $\sum_{i=1}^{k} a_i = 1$. Then,

$$m := \min\{P_i\}_{i=1}^{k} \leq \sum_{i=1}^{k} a_i P_i.$$

Proof: Since $P_i \geq m$ for each P_i. Now,

$$\sum_{i=1}^{k} a_i P_i \geq \sum_{i=1}^{k} a_i m = m \sum_{i=1}^{k} a_i = m. \ \square$$

Lemma 4.2: Let $a \in \mathbb{C}$. To each $z_1 \in \mathbb{C}$, there exists a homologous z such that

$$|z^2 - a|^2 \leq \frac{7}{16} + |a|^2,$$

with equality only in the cases where

$$\left[a = 3/4 \text{ and } z_1 \text{ homologous to } \frac{i}{2}\right],$$

or

$$\left[a = -3/4 \text{ and } z_1 \text{ homologous to } \frac{1}{2}\right].$$

Definition: We say that $z, z_1 \in \mathbb{C}$ are homologous if and only if $z - z_1 \in \mathbb{Z}[i]$.

Proof: Let $a = \alpha + ib$.

Case 1: $(\alpha \geq 0)$

To each z_2, there exist infinitey many homologous $z = x + yi$ such that $-\frac{1}{2} < y < \frac{1}{2}$, and from these choose one satisfying

$$-\frac{1}{2} + \left(\frac{1}{4} - y^2\right)^{1/2} < x \leq \frac{1}{2} + \left(\frac{1}{4} - y^2\right)^{1/2}.$$

Define functions:

$$P = |z^2 - a|^2 - |a|^2, \qquad Q = |(z-1)^2 - a|^2 - |a|^2, \qquad R = |(z+1)^2 - a|^2 - |a|^2.$$

Now,

$$\begin{aligned} P &= |z^2 - a|^2 - |a|^2 \\ &= (z^2 - a)(\overline{z}^2 - \overline{a}) - a\overline{a} \\ &= ((x+iy)^2 - (\alpha + i\beta))((x-iy)^2 - (\alpha - i\beta)) - (\alpha + i\beta)(\alpha - i\beta) \\ &= x^4 + 2x^2y^2 + y^4 - 2(x^2 - y^2)\alpha - 4xy\beta. \end{aligned}$$

Hence,

$$(1-x)P + xQ = u - 3u^2 + 2uv + v^2 + 2(v-u)\alpha, \text{ where } u = x - x^2 \text{ and } v = y^2.$$

Similarly,

$$(1-x^2-y^2)P+\frac{1}{2}(x^2+y^2+x)Q+\frac{1}{2}(x^2+y^2-x)R = y^2-3x^2+3y^4+6y^2x^2+3x^4.$$

Looking at subintervals of the original interval, consider:

$$\frac{1}{2}-\left(\frac{1}{4}-y^2\right)^{1/2} < x < \left(\frac{1}{4}-y^2\right)^{1/2}+\frac{1}{2}.$$

Now,

$$0 \le \left(x-\frac{1}{2}\right)^2 \le \frac{1}{4}-y^2,$$

and so $x-x^2 \le \frac{1}{4}$.

Similarly,

$$0 \le y^2 \le x-x^2 \le \frac{1}{4},$$

which gives $0 \le v \le u \le \frac{1}{4}$.

In this case

$$\begin{aligned}\min(P,Q) &\le (1-x)P+xQ = u-3u^2+2uv+v^2+2(u-v)\alpha\\ &\le u-3u^2+2u^2+u^2\\ &= u\\ &\le \frac{1}{4}\\ &< \frac{7}{16}.\end{aligned}$$

Consider the subinterval on the other side

$$-\frac{1}{2}+\left(\frac{1}{4}-y^2\right)^{1/2} < x \le \frac{1}{2}-\left(\frac{1}{4}-y^2\right)^{1/2}.$$

Now,

$$|x| \le \frac{1}{2} \qquad \text{and} \qquad \left(\frac{1}{2}\pm x\right)^2 \ge \frac{1}{4}-y^2, \qquad \text{hence} \quad x^2+y^2\pm x \ge 0.$$

Also,

$$1-x^2-y^2 \ge 0, \text{ since } |x| \le \frac{1}{2} \text{ and } |y| \le \frac{1}{2}.$$

Combining all these facts, $\min(P,Q,R) \le \frac{7}{16}-\frac{3}{2}x^2(1-2x^2)$. We have possible equality here only if $x=0$ and $z=x+iy$ and $z=iy$ and $y=\frac{1}{2}$.

[Missing remainder of talk.]

Appendix D

Jay Pantone: History of π

Some Early Estimates for π:

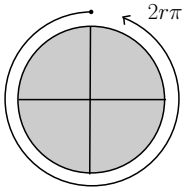
Perhaps the earliest estimate for π came from the Old Testament (translated around 550 BC):

> *And he made a molten sea, ten cubits from the one brim to the other: it was round all about, and his height was five cubits: and a line of thirty cubits did compass it round about.*
> - I Kings 7:23

which effectively estimates π as 3.

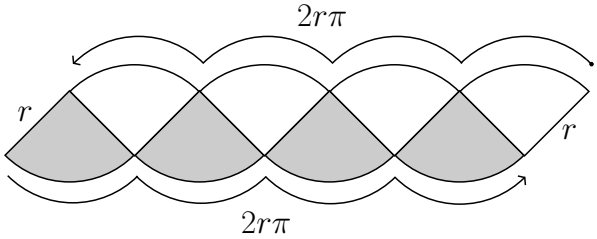
In fact, as early as 2000 BC, the Babylonians and the Egyptians already had reasonable estimates for π. While the method of estimation isn't known, they probably arose through construction of a circle with a fixed diameter, followed by measuring the circumference as accurately as possible. The Babylonians used the value $\pi = 3\frac{1}{8} = 3.125$, which was found on a tablet excavated in 1936 about 200 miles from Babylon. Meanwhile, the Egyptians used the value $\pi = \left(\frac{16}{9}\right)^2 = \frac{256}{81} \approx 3.1605\ldots$, which was written on the famous Ahmes Papyrus and discovered in an abandoned building in 1858.

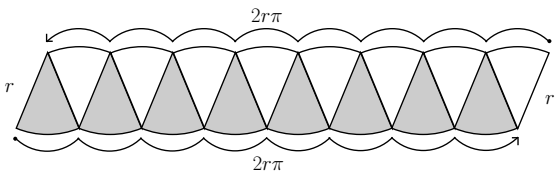
Despite only having estimations for the value of π, the ancients knew the formula for the area of a circle to be πr^2. Here is the most probable way they discovered this:



The circumference of a circle with radius r is $2r\pi$ by the definition of the constant π. We start with such a circle.

We take a second identical circle, split each up into four parts, and align them as shown. The length along the top and bottom is the circumference of the original circle.



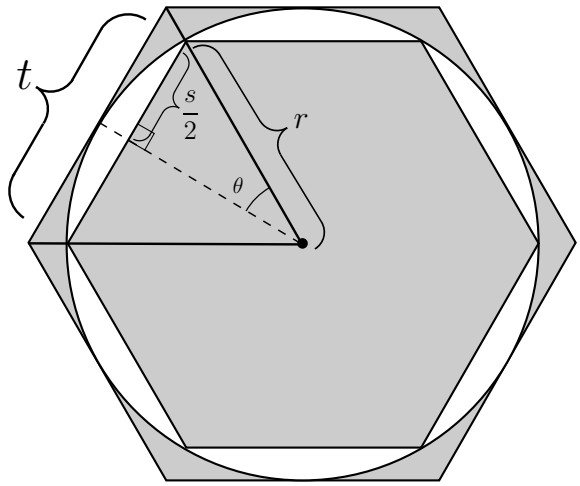


If we instead split each circle into eight equal parts and align them as shown, then the length along the top and bottom is still the circumference of the original circle, but the shape is closer to a rectangle.

Splitting the circle up into more and more parts, the resulting shape is closer and closer to a rectangle. This implies that the area of the two circles is $r \cdot 2r\pi = 2r^2\pi$, and hence the area of each circle is πr^2.

In the search for π's true value, **Archimedes** is thought to be the first to apply purely mathematical methods. In *On the Measurement of the Circle* he uses the following geometric construction to obtain bounds for π:

Construction by the Archimedean Method starts with a circle along with both an inscribed and circumscribed hexagon.



Observe that $s = 2r\sin\theta$ and $t = 2r\tan\theta$. Using $2\pi r$ as the circumference of the circle, and noting that the circumference of the circle lies between the circumferences of the inscribed and circumscribed polygons, we now have that

$$2rn\sin\theta < 2\pi r < 2rn\tan\theta, \qquad \text{and hence}$$
$$n\sin\theta < \pi < n\tan\theta.$$

In the hexagonal case, as above, we have $n = 6$ and $\theta = \frac{\pi}{6}$. As we double the number of sides k times, we have that:

$$2^k n\sin\frac{\theta}{2^k} < \pi < 2^k n\tan\frac{\theta}{2^k}.$$

As k is increases toward infinity, the bounds get closer to the actual value of π because the difference in circumference of the inscribed and circumscribed polygons gets closer to zero. **Archimedes** and others knew that for the starting hexagon, we have $\sin\theta = \frac{1}{2}$ and $\tan\theta = \frac{1}{\sqrt{3}}$. To increase k, they used various versions of the half-angle formulas.

For a millennium and a half, the Archimedean Method of estimating π reigned supreme. All that was needed to reach better and better estimates was the time and dedication to carry the computation further and further. Some of the estimates during this era are:

(380 AD) One of the Indian Siddhantas gives the value of π as $3\frac{177}{1250} = 3.1416$.

(450 AD) Chinese Mathematician **Tsu Chung-Chih** and his son **Tsu Keng-Chih** found

$$3.1415926 < \pi < 3.1415927.$$

It is likely that the Chinese were able to obtain such a high accuracy because they're knowledge of the digit 0 made them better equipped for calculation. An estimate of this level of accuracy would not be discovered in Europe for over 1000 years.

(598 AD) The Hindu mathematician **Brahmagupta** used $\pi = \sqrt{10} \approx 3.1623\ldots$. He likely observed that the perimeters of polygons with 12, 24, 48, and 96 sides inscribed in a circle of diameter 10, are given by the sequence $\sqrt{965}, \sqrt{981}, \sqrt{986}, \sqrt{987}$, and assumed that the sequence approached $\sqrt{1000} = 10\sqrt{10}$.

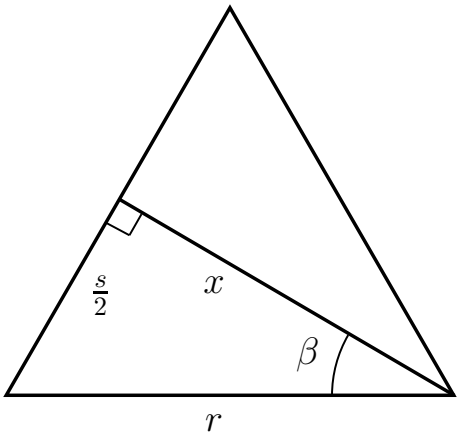
(1220 AD) **Leonardo of Pisa** - aka Fibonacci (literally, *Son of Bonaccio*) - used his era's new decimal arithmetic to obtain an estimation $\pi = \frac{846}{275} = 3.14\overline{18}$.

The New Age of π

François Viète:

After centuries of repeated applications of the Archimedean Method, the 16$^{\text{th}}$ century finally marked new progress. French mathematician **François Viète** (1540-1603) was primarily a lawyer, but did his most important mathematical work while in exile for his suspected sympathies toward the Protestant cause. It was in his famous *Analytical Art* that he coined some words we still use today, such as "negative", "coefficient", and "analytical". **Viète** used a method similar to the following to find the first ever analytical expression of π as a sequence of algebraic operations.

In the spirit of the Archimedean Method, **Viète** compared an n-gon to a $2n$-gon, but rather than comparing their perimeters, he compared their areas. Let the below diagram be one triangular section of the n-gon of radius r. Let 2β be the angle formed by this section, so that $\beta = \pi/n$.



Observe that the area of this triangle equals $\frac{1}{2}sx = (r\sin\beta)(r\cos\beta) = r^2\sin\beta\cos\beta$. Hence, the area of the n-gon is $nr^2\sin\beta\cos\beta$. Similarly, the area of the $2n$-gon is $2nr^2\sin\frac{\beta}{2}\cos\frac{\beta}{2} = nr^2\sin\beta$.

Hence, denoting $A(k)$ as the area of the k-gon, we have that:

$$\frac{A(n)}{A(2n)} = \frac{nr^2 \sin\beta\cos\beta}{nr^2\sin\beta} = \cos\beta.$$

Similarly,

$$\frac{A(n)}{A(4n)} = \frac{A(n)}{A(2n)}\frac{A(2n)}{A(4n)} = \cos\beta\cos\frac{\beta}{2}.$$

Repeating this process:

$$\frac{A(n)}{A(2^k n)} = \cos\beta\cos\frac{\beta}{2}\cdots\cos\frac{\beta}{2^k}.$$

Taking limits:

$$\lim_{k\to\infty}\frac{A(n)}{A(2^k n)} = \cos\beta\cos\frac{\beta}{2}\cos\frac{\beta}{4}\cos\frac{\beta}{8}\cdots.$$

Since $A(2^k n)$ approaches the area of the circle with radius r as $k\to\infty$, we have that

$$\frac{nr^2\cos\beta\sin\beta}{\pi r^2} = \cos\beta\cos\frac{\beta}{2}\cos\frac{\beta}{4}\cos\frac{\beta}{8}\cdots.$$

Hence

$$\pi = \frac{n\cos\beta\sin\beta}{\prod_{k=0}^{\infty}\cos\frac{\beta}{2^k}}.$$

Viète started with a square, setting $n=4$ and $\beta=\pi/4$. Applying a half-angle formula

$$\cos\frac{\theta}{2} = \sqrt{\frac{1}{2}+\frac{1}{2}\cos\theta},$$

our formula above yields

$$\pi = 2\cdot\frac{1}{\sqrt{\frac{1}{2}}\cdot\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}\cdot\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}}\cdots}.$$

This convergence was proved formally by **Ferdinand Rudio** (1856-1929) in 1891.

This formula converges slowly, and so it isn't useful for calculating digits. **Viète** himself used the Archimedean Method to calculate π to 9 digits. Not many mathematicians wasted their time with the Archimedean Method after this calculation, principally because it served no practical purpose. Knowing π to 40 digits provides enough accuracy to compute the circumference of the Milky Way to within the radius of a proton!

John Wallis:

In 1655, British Mathematician **John Wallis** (1616-1703), considered the area under a circular arc - for which he had a formula due to **Descartes** - and used very tedious methods to derive

$$\pi = 2\cdot\frac{2\cdot2\cdot4\cdot4\cdot6\cdot6\cdot\cdots}{3\cdot3\cdot5\cdot5\cdot7\cdot7\cdot\cdots}.$$

This was the first expression of π to have only rational quantities.

Lord Brouncker:

As discussed in class, British Mathematician **Lord William Brouncker** (1620-1684) used

Wallis' formula to compute a general continued fraction for π:

$$\frac{\pi}{4} = \cfrac{1}{1+\cfrac{1^2}{2+\cfrac{3^2}{2+\cfrac{5^2}{2+\cfrac{7^2}{2+\ddots}}}}}.$$

James Gregory:

The Scottish mathematician and astronomer **James Gregory** (1638-1675) made the next giant leap in the history of π. In 1671, he proved the equivalent of:

$$\arctan(x) = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots.$$

Setting $x = 1$, we have the formula

$$\pi = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right).$$

Though convergence is very slow (it takes over 300 terms to get two decimal places of accuracy), this is recognized as the first infinite series representation of π.

Isaac Newton:

The well-known British physicist, mathematician, astronomer, philosopher, and alchemist **Isaac Newton** could not resist spending some time working with π. He was aware of the Gregeory series, and knew that it converged far too slowly to be useful. Instead, he used the following derivation:

Through his study of "fluxions" (derivatives) and "Flowing Quantities" (integrals), **Newton** knew the equivalent of

$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin(x).$$

Using his own generalized binomial theorem, he also knew that

$$\int \frac{dx}{\sqrt{1-x^2}} = \int \left(1 + \frac{1}{2}x^2 + \frac{1\cdot 3}{2\cdot 4}x^4 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}x^6 + \cdots\right) dx.$$

Setting these two equal and integrating term-by-term, **Newton** found:

$$\arcsin(x) = x + \frac{1}{2}\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{x^5}{5} + \cdots.$$

Substituting $x = \frac{1}{2}$:

$$\pi = 6\left(\frac{1}{2} + \frac{1}{2\cdot 3\cdot 2^2} + \frac{1\cdot 3}{2\cdot 4\cdot 5\cdot 2^5} + \cdots\right).$$

This is another infinite series for π. However, it converges incredibly faster than the Gregory series. **Newton** himself used this method to calculate 15 correct decimal places for π, but later admitted in a letter "I am ashamed to tell you to how many figures I carried these computations, having no other business at the time."

John Machin:

Though he was a Professor of Astronomy at Gresham College in London, **John Machin** (1680-1750) is best known for his work with π, which came in 1706.

Let β be such that $\tan(\beta) = 1/5$. By the double angle formula:

$$\tan(2\beta) = \frac{2\tan(\beta)}{1-\tan^2(\beta)} = \frac{2/5}{24/25} = \frac{5}{12}.$$

$$\tan(4\beta) = \frac{2\tan(2\beta)}{1-\tan^2(2\beta)} = \frac{5/6}{119/144} = \frac{120}{119}.$$

We now notice that $\frac{120}{119}$ is just $\frac{1}{119}$ away from 1, which is $\tan\left(\frac{\pi}{4}\right)$. Using the formula for subtraction of angles inside of tangent:

$$\tan\left(4\beta - \frac{\pi}{4}\right) = \frac{\tan(4\beta) - \tan\left(\frac{\pi}{4}\right)}{1+\tan(4\beta)\tan\left(\frac{\pi}{4}\right)} = \frac{\tan(4\beta)-1}{1+\tan(4\beta)} = \frac{1/119}{239/119} = \frac{1}{239}.$$

From this, we find the formula:

$$\arctan(1/239) = 4\beta - \frac{\pi}{4} = 4\arctan(1/5) - \frac{\pi}{4}.$$

Hence,

$$\frac{\pi}{4} = 4\arctan(1/5) - \arctan(1/239),$$

and thus,

$$\frac{\pi}{4} = 4\left(\frac{1}{5} - \frac{1}{3\cdot 5^3} + \frac{1}{5\cdot 5^5} - \cdots\right) - \left(\frac{1}{239} - \frac{1}{3\cdot 239^3} + \frac{1}{5\cdot 239^5} - \cdots\right).$$

This expansion is so useful because the terms in the first infinite sum are easy to calculate and the terms in the second sum converge extremely fast. Machin used the formula soon after discovering it to calculate π to 100 decimal places. In the very same year, the symbol "π" was first used by **William Jones** to denote the "periphery" of a circle of diameter 1.

Leonhard Euler:

It goes without saying that **Leonhard Euler** (1707 - 1783) found many formulas for π. It was one of these formulas that helped him make his name as a mathematician among the rest of Europe's mathematicians. At the age of 28, **Euler** solved the famous *Basel Problem*: finding the sum of the squares of the reciprocals of the natural numbers. Here's how he did it:

Euler knew the formula

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots .$$

Solving for the equation $\sin(x) = 0$, we have roots $x = 0,\ \pm\pi,\ \pm 2\pi,\ \cdots$. So, making the assumption that we can treat this infinite polynomial like a finite polynomial, we can write $\sin(x)/x$ as:

$$\frac{\sin(x)}{x} = \left(1 + \frac{x}{\pi}\right)\left(1 - \frac{x}{\pi}\right)\left(1 + \frac{x}{2\pi}\right)\left(1 - \frac{x}{2\pi}\right)\cdots .$$

Multiplying this out and collecting all terms with the coefficient x^2, we find that the x^2 term of $\sin(x)/x$ has coefficient:

$$-\left(\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \cdots\right) = -\frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

Since the coefficient of x^2 in $\sin(x)/x$ is $-1/3! = -1/6$ as shown in the first formula (before factoring out an x), we conclude that:

$$-\frac{1}{6} = -\frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}, \quad \text{and hence:} \quad \frac{\pi^2}{6} = \sum_{n=1}^{\infty}\frac{1}{n^2}.$$

Johann Zacharias Dase:

Johann Dase (1824-1861) is the most well-known example of a "human calculator". Though he was completely unable to comprehend even basic mathematical theory, he could calculate large quantities in his head extremely rapidly. He used a formula derived similarly to Machin's Formula

$$\frac{\pi}{4} = \arctan\left(\frac{1}{2}\right) + \arctan\left(\frac{1}{5}\right) + \arctan\left(\frac{1}{8}\right)$$

to calculate 200 digits of π in his head, over the span of two months. After going on tour around Europe showcasing his ability, he calculated the log of the first million numbers to seven decimal places, created a table of the hyperbolic functions, and was halfway through factoring the numbers 7,000,000 - 10,000,000 when he died.

Johann Heinrich Lambert:

Despite all the attention π received over thousands years, it wasn't until 1767 it was proved that π was irrational, by **Johann Heinrich Lambert** (1728-1777). His theorem actually proved that if x is rational and nonzero, then $\tan(x)$ is irrational. Since $\tan\left(\frac{\pi}{4}\right) = 1$, which is rational, it follows that $\frac{\pi}{4}$ and hence π are irrational. In 1794, **Adrien-Marie Legendre** (1752-1833) made the proof more rigorous, and additionally proved that π^2 is irrational.

Carl Louis Ferdinand von Lindemann:

In 1882, **von Lindemann** (1852-1939) proved that π is a transcendental number, i.e., it is not the root of any polynomial with integer coefficients. **Lindemann** extended ideas from **Charles Hermite** (1822-1901) so that **Euler's Equation** $e^{i\pi} + 1 = 0$ implied the transcendence of $i\pi$. Since i is algebraic, this implied the transcendence of π.

Circle Squarers:

The proof of the transcendence of π finally shut the book on the ancient problem of "squaring the circle". The challenge posed by the ancient Greeks was to construct a square with the same area as a given circle in a finite number of steps using only a compass and ruler. It is shown in a typical graduate-level algebra class that any constructible lengths under these restraints must be algebraic of

degree a power of two. If it was possible to square the circle, then the length π would be constructible. Since π is not algebraic, it is not possible to circle the square.

However, the fact that an act is impossible has never stopped those seeking fame and glory to try it, and this is no different. Even to this day, mathematics departments occasionally get proofs from those known as "circle-squarers" or "cyclometers" who claim to have successfully squared the circle.

In 1874, **John A. Parker** published The Quadrature of the Circle, in which he "proved" the value $\pi = \frac{20612}{6561} \approx 3.14159427\ldots$. This is closest to a value used by **Valentinus Otho** in 1573. The argument set forth by **Parker** amounted to the claim that the circumference should be measured as the length surrounding the outside of the circle, rather than the length of the boundary of the circle.

In 1897, the **Indiana State Legislature** almost passed a bill legislating an incorrect value of π. The construction of π and other geometric claims was a gift from **Edwin Goodwin**, a physician. In fact, the set of constructions gave two different values of π - the first was $\frac{4}{5/4} = 3.2$. The second, though likely accidental, was $\frac{16}{\sqrt{3}} \approx 9.2376\ldots$. Dr. Goodwin promised that with the passing of the bill, the state of Indiana would have free rights to the knowledge, while other states would have to pay royalties to publish his findings. The bill was directed to the *Committee on Swamp Lands* (no justification given in the legislative records), who then passed it on to the *Committee on Education*, which recommended its passage. The bill immediately passed the House unanimously and went to the Senate. Thankfully, on the day of the vote **Professor C.A. Waldo** was visiting the legislature on behalf of the Department of Mathematics of Purdue University on separate business. Upon hearing the bill read, he quickly intervened, giving the Senators a quick math lesson. The bill was tabled, and has not been revisited since.

Perhaps the most egregious example of a circle-squarer is **Carl Theodore Heisel**, who published The Circle Squared Beyond Refutation, in which he not only claims to square the circle, but also proves that decimals are inexact, disproves Pythagorean's Theorem as a special case, and shows the square roots of the first 100 positive integers to be rational numbers. His claim is $\pi = 3\frac{13}{81} = \frac{256}{81}$, which is the same value used in the Ahmes Papyrus over 4000 years ago. **Heisel** paid to have thousands of copies printed, and he distributed copies of his work to libraries and universities across the country, to aide in the education of mathematical students. **Heisel** did not think highly of mathematicians, saying in his preface:

> *As a class modern mathematicians do not seem to possess any doubt. They do not question anything. They blindly accept what is taught in the books as absolute and final. They learn their mathematics as a parrot learns language by simple imitation.*

The Computer Era

From nearly the instant that the first computer was turned on in the 1950s, programmers and computer enthusiasts have competed over who can calculate the most digits of π at the fastest speed. They do however owe a great deal to the mathematicians that continued to develop more efficient means of computing π. Many of the below algorithms use techniques such as the Fast Fourier Transform to help speed up arithmetic. The popular algorithm through the 1970s involved Machin's Formula at a high precision. Although this was not remarkably efficient, it was the best algorithm of the time.

Srinivasa Ramanujan:

Although **Srinivasa Ramanujan** (1887-1920) was not alive to see the computer era, much of his work was not discovered until this time period. **Ramanujan** had many absolutely astounding formulas for π. One of the most famous of these, discovered around 1910, is:

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}}.$$

This equation was used by American mathematician and programmer **Bill Gosper** (1943-) to calculate 17 million digits of π in 1985. The use of this equation yields an average of eight additional correct digits for each iteration of the sum. While this was certainly an advance, it was still only a linear algorithm.

Arithmetic-Geometric Mean:

The Arithmetic-Geometric Mean (AGM) iteration - made popular by **Gauss** (1777-1855) - provides a treasure trove of algorithmic methods to compute irrational numbers. In this section we develop the AGM, and in the next section we'll derive the Salamin-Brent Algorithm for computing digits of π with quadratic convergence.

Between two numbers x and y, the arithmetic mean is $\dfrac{x+y}{2}$ and the geometric mean is $\sqrt{xy}$. To find the Arithmetic-Geometric Mean of x and y, denoted $\mathrm{AGM}(x, y)$, we first set

$$a_0 := \max(x, y), \qquad b_0 := \min(x, y).$$

Now, we compute the sequences

$$a_{n+1} := \frac{a_n + b_n}{2}, \qquad b_{n+1} := \sqrt{a_n b_n}.$$

Remarkably, the sequences $\{a_n\}$ and $\{b_n\}$ not only converge but converge to the same limit - we call this limit the Arithmetic-Geometric Mean of a_0 and b_0. To see this, first recall that the geometric mean of two distinct numbers is always strictly less than their arithmetic mean. So:

$$b_{n+1} = \sqrt{a_n b_n} \geq \sqrt{b_n^2} = b_n.$$

Therefore, the sequence of geometric means is increasing. It is also bounded above by a_0, and therefore the sequence converges. Define $\mathrm{AGM}(a_0, b_0) := \lim_{n\to\infty} b_n$. Now, note that

$$\frac{b_{n+1}^2}{b_n} = \frac{\sqrt{a_n b_n}^2}{b_n} = \frac{a_n b_n}{b_n} = a_n.$$

Hence:

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{b_{n+1}^2}{b_n} = \frac{\left(\lim_{n\to\infty} b_{n+1}\right)^2}{\lim_{n\to\infty} b_n} = \frac{\left(\mathrm{AGM}(a_0,b_0)\right)^2}{\mathrm{AGM}(a_0,b_0)} = \mathrm{AGM}(a_0,b_0).$$

Thus, the sequences $\{a_n\}$ and $\{b_n\}$ both converge to the same limit. Since

$$a_{n+1} = \frac{a_n + b_n}{2} < \frac{a_n + a_n}{2} = a_n,$$

we have that

$$b_0 \le b_1 \le b_2 \le \cdots \le \mathrm{AGM}(a_0,b_0) \le \cdots \le a_2 \le a_1 \le a_0.$$

Salamin-Brent Algorithm:

In 1976, **Eugene Salamin** (no dates found) and **Richard Brent** (1946-) independently found an algorithm using the Arithmetic-Geometric Mean that could *quadratically* produce correct digits of π. Here is that algorithm:

Set $a_0 = 1$, $b_0 = \frac{1}{\sqrt{2}}$, and $s_0 = \frac{1}{2}$. For $k = 1, 2, 3, \ldots$, compute

$$a_k = \frac{a_{k-1} + b_{k-1}}{2},$$

$$b_k = \sqrt{a_{k-1} b_{k-1}},$$

$$c_k^2 = a_k^2 - b_k^2,$$

$$s_k = s_{k-1} - 2^k c_k^2,$$

$$p_k = \frac{2a_k^2}{s_k}.$$

Then, p_k converges quadratically to π. Successive iterations of this algorithm give $1, 4, 9, 20, 42, 85, 173, \ldots$ correct digits.

Of course, it is true that this algorithm requires the ability to perform high-precision square root operations. However, using a method called Newton Iteration, we can perform a square root in about the time it takes to perform three multiplications, making this algorithm must faster than the previous linear methods.

Below is the proof that the p_k sequence in the Salamin-Brent algorithm does converge to π. The fact that the convergence is quadratic can be found in **Salamin's** paper *Computation of π Using Arithmetic-Geometric Mean.*

Consider the elliptic integrals:

$$K(k) = \int_0^{\pi/2} \left(1 - k^2 \sin^2(t)\right)^{-1/2} dt,$$

$$E(k) = \int_0^{\pi/2} \left(1 - k^2 \sin^2(t)\right)^{1/2} dt.$$

If there exists k' such that $k^2 + k'^2 = 1$, then we can define two more elliptic integrals:

$$K'(k) := K(k') = K\left(\sqrt{1-k^2}\right),$$

$$E'(k) := E(k') = E\left(\sqrt{1-k^2}\right).$$

Now, we define symmetric versions of these last two:

$$\begin{aligned} I(a,b) &= \int_0^{\pi/2} \left(a^2\cos^2(t) + b^2\sin^2(t)\right)^{-1/2} dt \\ &= \frac{1}{a}\int_0^{\pi/2} \left(\cos^2(t) + \left(\frac{b}{a}\right)^2 \sin^2(t)\right)^{-1/2} dt \\ &= \frac{1}{a}\int_0^{\pi/2} \left(1 - \sin^2(t) + \left(\frac{b}{a}\right)^2 \sin^2(t)\right)^{-1/2} dt \\ &= \frac{1}{a}\int_0^{\pi/2} \left(1 - \left(1 - \frac{b^2}{a^2}\right)\sin^2(t)\right)^{-1/2} dt \\ &= a^{-1}K'(b/a). \end{aligned}$$

$$\begin{aligned} J(a,b) &= \int_0^{\pi/2} \left(a^2\cos^2(t) + b^2\sin^2(t)\right)^{1/2} dt \\ &= a\int_0^{\pi/2} \left(\cos^2(t) + \left(\frac{b}{a}\right)^2 \sin^2(t)\right)^{1/2} dt \\ &= a\int_0^{\pi/2} \left(1 - \sin^2(t) + \left(\frac{b}{a}\right)^2 \sin^2(t)\right)^{1/2} dt \\ &= a\int_0^{\pi/2} \left(1 - \left(1 - \frac{b^2}{a^2}\right)\sin^2(t)\right)^{1/2} dt \\ &= aE'(b/a). \end{aligned}$$

The transformations given by English mathematician **John Landen** (1719-1790) yield:

$$I(a_n, b_n) = I(a_{n+1}, b_{n+1}),$$

$$J(a_n, b_n) = 2J(a_{n+1}, b_{n+1}) - a_n b_n I(a_{n+1}, b_{n+1}).$$

So, letting $M := \mathrm{AGM}(a_0, b_0)$,

$$I(a_0, b_0) = I(M, M) = \frac{1}{M} = \frac{K'(1)}{M} = \frac{K(0)}{M} = \frac{\pi}{2M}.$$

In the *Handbook of Mathematical Functions*, by **Abramowitz** and **Stegun**, they derive

$$J(a_0, b_0) = \left(a_0^2 - \frac{1}{2}\sum_{j=0}^{\infty} 2^j c_j^2\right) I(a_0, b_0),$$

where $c_n^2 = a_n^2 - b_n^2$.

The **Legendre** Relation of elliptic integrals tells us that

$$K(k)E'(k) + K'(k)E(k) - K(k)K'(k) = \frac{\pi}{2}.$$

Making substitutions and multiplying through by aa', we have:

$$a^2 I(a,b)J(a',b') + (a')^2 I(a',b')J(a,b) - a^2(a')^2 I(a,b)I(a',b') = aa'\frac{\pi}{2}.$$

Now, set $a = a' = 1$ and $b = b' = 1/\sqrt{2}$. Using the relations above, and defining $M := \mathrm{AGM}(1, 1/\sqrt{2})$, we see:

$$\frac{\pi}{2} = 2\left(\frac{\pi}{2M}\right)\left(\left(1 - \frac{1}{2}\sum_{j=0}^{\infty} 2^j c_j^2\right)\left(\frac{\pi}{2M}\right)\right) - \frac{\pi^2}{4M^2}$$

$$\frac{\pi}{2} = \left(\frac{\pi^2}{4M^2}\right)\left(2\left(1 - \frac{1}{2}\sum_{j=0}^{\infty} 2^j c_j^2\right) - 1\right)$$

$$\frac{2M^2}{\pi} = 2\left(1 - \frac{1}{2}\sum_{j=0}^{\infty} 2^j c_j^2\right) - 1$$

$$\pi = \frac{2M^2}{2\left(1 - \frac{1}{2}\sum_{j=0}^{\infty} 2^j c_j^2\right) - 1} = \frac{2M^2}{1 - \sum_{j=0}^{\infty} 2^j c_j^2} = \frac{2M^2}{\frac{1}{2} - \sum_{j=1}^{\infty} 2^j c_j^2}.$$

This formula matches the p_k terms of the algorithm as $k \to \infty$. Hence, the sequence $\{p_k\}$ converges to π. $\square$

Jonathan and Peter Borwein:

The brothers **Jonathan Borwein** (1951-) and **Peter Borwein** (1953-) together in 1985 found methods similar to the Salamin-Brent Algorithm, but with faster convergence. The most notable version is the quartic:

Set $a_0 = 6 - 4\sqrt{2}$ and $y_0 = \sqrt{2} - 1$. For $k = 1, 2, 3, \ldots$, compute

$$y_{k+1} = \frac{1 - (1 - y_k^4)^{1/4}}{1 + (1 - y_k^4)^{1/4}},$$
$$a_{k+1} = a_k(1 + y_{k+1})^4 - 2^{2k+3} y_{k+1}(1 + y_{k+1} + y_{k+1}^2).$$

Then, p_k converges *quartically* to π.

The quartic algorithm was used by **Yasumasa Kanada** (no dates found) of the University of Tokyo to set the at-the-time record of 6.4 billion digits. He would later use this method to calculate up to 5 trillion digits.

It has since been shown that there are m^{th} order approximations for all m. In the paper Approximations to π via the Dedekind eta function (1996) by **J. Borwein** and **F. Garvan**, they give the following example of a nonic algorithm:

Set $a_0 = \frac{1}{3}$, $r_0 = \frac{(\sqrt{3}-1)}{2}$, and $s_0 = (1 - r_0^3)^{1/3}$. For $k = 1, 2, 3, \ldots$, compute

$$t = 1 + 2r_k$$

$$u = [9r_k(1 + r_k + r_k^2)]^{1/3}$$

$$v = t^2 + tu + u^2$$

$$m = \frac{27(1 + s_k + s_k^2)}{v}$$

$$a_{k+1} = ma_k + 3^{2k-1}(1 - m)$$

$$s_{k+1} = \frac{(1 - r_k)^3}{(t + 2u)v}$$

$$r_{k+1} = (1 - s_k^3)^{1/3}$$

Now, $1/a_k$ converges *nonically* to π

It is worth noting that in terms of computer efficiency, the quartic algorithm is the most efficient of these. Though higher order algorithms converge faster, they are vastly more expensive in terms of computational time per stage.

The Chudnovsky Brothers:

In 1989, brothers **David Chudnovsky** (1947-) and **Gregory Chudnovsky** (1952-) discovered a formula for π in the spirit of the **Ramanujan** formula above:

$$\frac{1}{\pi} = 12 \sum_{k=0}^{\infty} \frac{(-1)^k (6k)!(13591409 + 545140134k)}{(3k)!(k!)^3 640320^{3k+3/2}}.$$

The **Chudnovsky Brothers** used this formula to calculate over 1 billion digits of π in the same year. Each iteration produces an average of 14 additional correct digits.

The current record of just over 10 trillion (decimal) digits, set in October 2011 by **Alexander Yee** and **Shigeru Kondo**, used this formula. The total computation time was 191 days on a very powerful desktop computer. The computation required 44 terabytes of disk space. (For comparison, it is often claimed that the Library of Congress contains roughly 10 terabytes of uncompressed print text.)

Rabinowitz-Wagon Algorithm:

In 1990, **Stanley Rabinowitz** (1947-) and **Stan Wagon** (1951-) developed a "spigot algorithm" for π. A "spigot algorithm" is one which can output the digits of π one at a time (in order only), but which does not use the previous digits as part of the computation of the new digits. However, an extraordinary amount of memory is needed to run the algorithm for a large number of digits. Because of this, it is never used for record-breaking attempts. It's attractiveness is in the fact that it uses only integer operations (additions, multiplication, reduction by a modulus), and therefore does not require high-precision floating point operations.

Computing Individual Digits of π:

After decades of finding and improving algorithms for computing π to high precision, all of which necessitated computing the first $d - 1$ digits of π in order to compute the d^{th} digit, it came as a great

surprise in 1996 that it was actually possible to compute individual hexadecimal (i.e., base 16) digits of π. This ability emerges from the amazing formula:

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i} \left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right).$$

The formula was found by **Simon Plouffe** (1956-) using an algorithm called *Pari-Gp* (slightly modified) that searches for integer relations for real numbers. This is similar to the more famous algorithm named *PSLQ*. After the formula was found, it was proved as follows:

It's clear that

$$\frac{x^{k-1}}{1-x^8} = \sum_{i=0}^{\infty} x^{k-1+8i}.$$

Now observe that

$$\int_0^{1/\sqrt{2}} \frac{x^{k-1}}{1-x^8} dx = \int_0^{1/\sqrt{2}} \sum_{i=0}^{\infty} x^{k-1+8i} dx = \sum_{i=0}^{\infty} \frac{x^{k+8i}}{8i+k} \Bigg|_0^{1/\sqrt{2}} = \frac{1}{2^{k/2}} \sum_{i=0}^{\infty} \frac{1}{16^i (8i+k)}.$$

Hence,

$$\sum_{i=0}^{\infty} \frac{1}{16^i} \left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right) = \int_0^{1/\sqrt{2}} \frac{4\sqrt{2} - 8x^3 - 4\sqrt{2}x^4 - 8x^5}{1-x^8}$$

Substituting $y := \sqrt{2}x$:

$$\int_0^{1/\sqrt{2}} \frac{4\sqrt{2} - 8x^3 - 4\sqrt{2}x^4 - 8x^5}{1-x^8} = 16 \int_0^1 \frac{4\sqrt{2} - 2\sqrt{2}y^3 - \sqrt{2}y^4 - \sqrt{2}y^5}{16-y^8} \frac{dy}{\sqrt{2}}$$

$$= 16 \int_0^1 \frac{4 - 2y^3 - y^4 - y^5}{16-y^8} dy = 16 \int_0^1 \frac{(y-1)(y^2+2)(y^2+2y+2)}{(y^2+2y+2)(y^2-2y+1)(y^2+2)(y^2-2)} dy$$

$$= 16 \int_0^1 \frac{y-1}{y^4-2y^3+4y-4} dy.$$

Using partial fraction decomposition:

$$\frac{16y-16}{y^4-2y^3+4y-4} = \frac{4y}{y^2-2} - \frac{4y-8}{y^2-2y+2}.$$

So,

$$16 \int_0^1 \frac{y-1}{y^4-2y^3+4y-4} dy = \int_0^1 \frac{4y}{y^2-2} dy - \int_0^1 \frac{4y-8}{y^2-2y+2} dy$$

$$= 2\ln(y^2-2)\big|_0^1 - 2\ln(y^2-2y+2)\big|_0^1 - 4\arctan(1-y)\big|_0^1$$

$$= 2(i\pi - (i\pi + \ln(2))) + 2\ln(2) - 4\left(0 - \frac{\pi}{4}\right) = \pi. \ \square$$

The key value in this formula is the "16^i". This allows us to use the following method to compute the d^{th} hexadecimal digit of π:

Let

$$S := \sum_{i=0}^{\infty} \frac{1}{16^i} \left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right).$$

To find the d^{th} hexadecimal digit of S, we can look at the first hexadecimal digit of the fractional part of $16^{d-1}S$ which will be denoted by $\text{frac}(16^{d-1}S)$.
Define

$$S_1 := \sum_{k=0}^{\infty} \frac{1}{16^k(8k+1)}, \qquad S_2 := \sum_{k=0}^{\infty} \frac{1}{16^k(8k+4)},$$
$$S_3 := \sum_{k=0}^{\infty} \frac{1}{16^k(8k+5)}, \qquad S_4 := \sum_{k=0}^{\infty} \frac{1}{16^k(8k+6)}.$$

so that

$$S = 4S_1 - 2S_2 - S_3 - S_4.$$

Note that for S_1:

$$\text{frac}(16^{d-1}S_1) = \text{frac}\left(\sum_{k=0}^{\infty} \frac{16^{d-1-k}}{8k+1} \right) = \text{frac}\left(\sum_{k=0}^{d-1} \frac{16^{d-1-k}}{8k+1} + \sum_{k=d}^{\infty} \frac{16^{d-1-k}}{8k+1} \right)$$

$$= \left[\sum_{k=0}^{d-1} \frac{16^{d-1-k} \bmod 8k+1}{8k+1} \bmod 1 \;+\; \sum_{k=d}^{\infty} \frac{16^{d-1-k}}{8k+1} \bmod 1 \right] \bmod 1.$$

A computer can very rapidly calculate the term sum on the left via repeated squaring, modular reduction, and floating point division. Only a few terms need to be calculated of the sum on the right to prevent rounding errors. The result is a fraction between 0 and 1. Repeat this process with S_2, S_3, and S_4, then calculate S from these four quantities, reducing once more mod1 if needed. Then, after converting to hexadecimal, the first digit of the remaining fractional part is the d^{th} hexadecimal digit of π.

This scheme is not significantly faster at calculating all digits of π, but it does allow the calculation of any specific digit in a much shorter time. In 1997, **Fabrice Bellard** (1972-) improved the algorithm to yield a ~43% increase in efficiency. An employee of "Yahoo!" used their distributed computational abilities to calculate the two quadrillionth binary bit of π - which is 0. The calculation took 23 days on over 1000 individual machines running in parallel.

At the time that this algorithm was published, the authors had no similar scheme that could compute the decimal digits of π. They were not able to find any simple formula with "10" in the place of "16". However, later in the same year (1996), **Plouffe** devised a somewhat unrelated and more complicated method of calculating an arbitrary decimal digit of π using low memory, but in $O(n^3 \log^3(n))$ time, which is highly inefficient. Again, **Bellard** refined the technique to obtain an $O(n^2)$ algorithm. Additional refinements have been made in that time, though finding a specific decimal digit still remains less efficient than the elegant process described above of finding a specific hexadecimal digit.

Open Questions

Normality:

Despite thousands of years of investigation, there remain many unsolved questions about the famous constant π. It took most of the last four millenium to discover that π is both irrational and transcendental. The next logical step is to determine the normality of π, i.e. whether each n-length string of digits occurs with limiting frequency b^{-n} in all bases b. All calculations of digits of π seem to suggest normality. However, in addition to the open question of whether π is normal in any particular base, it hasn't even been shown that any particular digit repeats infinitely often!

The digit-extraction algorithms above have provided a new avenue for recent investigation. It is thought that the normality of π is equivalent to a "plausible" conjecture in the field of chaos theory.

Algebraic Independence:

Another famous open question is whether e and π are algebraically independent. It was shown that $\{\pi, e^\pi, \Gamma(\frac{1}{4})\}$ is an algebraically independent set over $\mathbb{Q}$ in 1996 by **Yuri Nesterenko** (1946-).

Irrationality Measure:

The irrationality measure of a real number gives a sense of how closely the number can be approximated by rationals. For example, algebraic numbers all have irrationality measure 2 (**Roth**, 1955). However, there are transcendental numbers that also have irrationality type 2: for example, the constant e.

The first irrationality measure of π was given by **Mahler** in 1953:

$$\left|\pi - \frac{p}{q}\right| > \frac{1}{q^{42}}, \quad q \geq 2.$$

In 1974, **Mignotte** improved this:

$$\left|\pi - \frac{p}{q}\right| > \frac{1}{q^{20.6}}, \quad q \geq 2.$$

The **Chudnovsky Brothers** mentioned above refined this to:

$$\left|\pi - \frac{p}{q}\right| > \frac{1}{q^{14.65}}, \quad q \text{ large}.$$

The current best measure was found by **Salikhov** in 2008:

$$\left|\pi - \frac{p}{q}\right| > \frac{1}{q^{7.6063\ldots}}, \quad q \text{ large}.$$

However, in April **Alekseyev** showed that if the Flint Hills series

$$\sum_{n=1}^{\infty} \frac{\csc^2(n)}{n^3}$$

converges, then this implies an irrationality measure

$$\left|\pi - \frac{p}{q}\right| > \frac{1}{q^{2.5}},$$

which is remarkably better than the current estimate. The Flint Hills series appears to converges upon visual inspection, but this convergence has not been proven. The difficulty is that $\csc^2(n)$ can sporadically take very large values.

When considering the intricate and vast history of π - from the Ancient Babylonians and Egyptians all the way through the computer age - it is not unthinkable that it could be hundreds if not thousands of years before these open questions are finally settled.

Appendix E

Meng Liu: Certain Trigonometric Values

Numbering follows "Irrational Numbers", by Niven.

Definition: Define

$$F_n(x) = \prod_{\substack{k=0 \\ \gcd(k,n)=1}}^{n-1} \left(x - e^{2\pi ik/n}\right). \tag{3.2}$$

Theorem 3.4: For $n \geq 1$, we have that

$$x^n - 1 = \prod_{d|n} F_d(x)$$

the product being over all divisors d of n.

Theorem 3.5: For $n \geq 1$, $F_n(x)$ is a monic polynomial of degree $\varphi(n)$ with rational integral coefficients.

Lemma 3.6: Let ω be a primitive n^{th} root of unity. Define $f(x)$ to be the minimal polynomial of ω. If p is any prime such that $\gcd(p, n) = 1$ and if ρ is any root of $f(x) = 0$, then ρ^p is also a root.

Theorem 3.7: The cyclotomic polynomial $F_n(x)$ is irreducible over the field of rational numbers.

Proof: (sketch) We know that $f(x)$ is monic and irreducible. If we can prove that any primitive n^{th} root of unity is a root of $f(x)$, then it follows that $f(x) = F_n(x)$. Any primitive n^{th} root of unity can be written as a power of a particular root ω, say ω^t, with $\gcd(t, n) = 1$. We can factor $t = p_1 p_2 \cdots p_s$ into (possibly not distinct) primes. Now, we have that ω^{p_1} is a primitive n^{th} root of unity, and hence so is $\omega^{p_1 p_2} = \omega^{p_1 p_2}$, etc, and hence so is ω^t.

By an extension of an earlier argument, we have that

$$x^{\varphi(n)} F_n(x^{-1}) = F_n(x).$$

Note that x^j and $x^{\varphi(n)-j}$ have the same coefficient. Also, for $n > 2$, we have that $\varphi(n) = 2m$ for some $m \in \mathbb{N}$. Now,

$$x^{-m} F_n(x) = (x^m + x^{-m}) + a_1(x^{m-1} + x^{1-m}) + \cdots + a_{m-1}(x + x^{-1}) + a_m.$$

By an identity

$$x^k + x^{-k} = (x + x^{-1})(x^{k-1} + x^{1-k}) - (x^{k-2} + x^{2-k}).$$

So, the polynomial $x^{-m}F_n(x)$ is a monic polynomial in $x+x^{-1}$, which we denote

$$\Psi(x+x^{-1}).$$

This completes the proof. $\square$

Lemma 3.8: Let $n > 2$ and let $m = \dfrac{\varphi(n)}{2}$. Then, $x^{-m}F_n(x)$ is a monic polynomial in $x+x^{-1}$, say $\Psi_n(x+x^{-1})$, with integral coefficients. Also, $\Psi_n(x)$ is irreducible of degree m.

Proof: Suppose toward a contradiction that $\Psi_n(x) = h_1(x)h_2(x)$, with $\deg(h_1), \deg(h_2) > 1$, then denote $r := \deg(h_1)$, and so $\deg(h_2) = m - r$. We have that

$$F_n(x) = x^m\Psi_n(x) = [x^r h_1(x)][x^{m-r}h_2(x)]$$

for all n. This contradicts **Theorem 3.7**. $\square$

Theorem 3.9: (D. H. Lehmer) If $n > 2$ and $\gcd(k,n) = 1$, then $2\cos\left(\dfrac{2\pi k}{n}\right)$ is an algebraic integer of degee $\dfrac{\varphi(n)}{2}$. For positive $n \neq 4$, we have that $2\sin\left(\dfrac{2\pi k}{n}\right)$ is an algebraic integers of degree $\varphi(n)$, $\dfrac{\varphi(n)}{4}$, or $\dfrac{\varphi(n)}{2}$, according to whether $\gcd(n,8) < 4$, $\gcd(n,8) = 4$, or $\gcd(n,8) > 4$.

Proof: Since $e^{2\pi ik/n}$ is a root of $F_n(x)$, it follows that

$$2\cos\left(\frac{2\pi k}{n}\right) = e^{\frac{2\pi ik}{n}} + e^{-\frac{2\pi ik}{n}}$$

is a root of $\Psi_n(x)$. Hence this number is an algebraic intger. By **Lemma 3.8**, the algebraic degree of this number is $\varphi(n)/2$.

Now, observe that

$$2\sin\left(\frac{2\pi k}{n}\right) = 2\cos\left(\frac{2\pi(4k-n)}{4n}\right).$$

We will now handle the different cases for

$$\frac{4k-n}{4n}.$$

This fraction is in its lowest terms if n is odd. Otherwise, it reduces to a fraction with a denominator $2n$ if $n \equiv 2 \pmod 4$ or it reduces to a fraction with denominator $\leq n$ if $n \equiv 0 \pmod 4$. Let d denote the degree of the algebraic integer

$$2\sin\left(\frac{2\pi k}{n}\right).$$

Case 1: (n odd)

We find that $d = \dfrac{\varphi(4n)}{2} = \varphi(n)$.

Case 2: ($n \equiv 2 \pmod 4$)

We find that $d = \dfrac{\varphi(2n)}{2} = \varphi(n)$.

Case 3 ($n \equiv 0 \pmod 4$)

Subcase 3.1: ($n \equiv 0 \pmod 8$)

Now, since k is odd (since n is even), we have that $d = \dfrac{\varphi(n)}{2}$.

Subcase 3.2: ($n \equiv 4 \pmod 8$)

Now, we have that the denominator of $\dfrac{4k-n}{4n}$ will be reduced in lowest terms to $n/4$. Hence we have $d = \dfrac{\varphi(n)}{4}$.

This covers all cases. □

Lemma 3.10: If $\mathbb{Q}$ is the field of rational numbers, then $\mathbb{Q}(\cos(2\theta))$ is a subfield of $\mathbb{Q}(\sin(\theta))$, $\mathbb{Q}(\cos(\theta))$, and $\mathbb{Q}(\tan(\theta))$, for all θ for which $\tan(\theta)$ exists.

Proof: Rewrite

$$\cos(2\theta) = 2\cos^2(\theta) - 2 = 1 - 2\sin^2(\theta) = \frac{2}{1+\tan^2(\theta)}.$$

Now, the subfield condition is clear. □

Theorem 3.11: For $n > 4$ and $\gcd(k, n) = 1$, the degree of $\tan\left(\dfrac{2\pi k}{n}\right)$ is $\varphi(n)$, $\dfrac{\varphi(n)}{2}$ or $\dfrac{\varphi(n)}{4}$, according to whether $\gcd(n, 8) < 4$, $\gcd(n, 8) = 4$, or $\gcd(n, 8) = 8$.

Appendix F

Ying Guo: Transcendence of $\sum_{n=0}^{\infty} \alpha^{2^n}$

Definition: Let $\alpha \in \mathbb{C}$ be algebraic of degree d and let $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_d$ be the conjugates of α, i.e. the roots of the minimal polynomial of α, which we denote $P_\alpha(x)$. We define the house of α to be the positive real number defined by

$$\overline{|\alpha|} := \max_{i \in [d]} |\alpha_i|.$$

Theorem 1: Let α and β be two algebraic numbers. Then,

$$\overline{|\alpha + \beta|} \leq \overline{|\alpha|} + \overline{|\beta|},$$
$$\overline{|\alpha\beta|} \leq \overline{|\alpha|}\,\overline{|\beta|}.$$

Theorem 2: Let α be a non-zero algebraic number of degree d. Then,

$$|\alpha| \geq \left(\overline{|\alpha|}\right)^{-d+1} \cdot (\mathrm{den}(\alpha))^{-d}$$

where $\mathrm{den}(\alpha)$ is the smallest integer by which we can multiply α to get an algebraic integer. This exists by a theorem from a previous class.

Lemma 1: Let $\Omega := \{x \in X \mid |x| < 1\}$. Then, the function

$$f(x) = \sum_{n=0}^{\infty} x^{2^n},$$

which is analytic in Ω, is transcendental.

Theorem 3: Let α be a nonzero algebraic integer with $|\alpha| < 1$. Then,

$$f(\alpha) = \sum_{n=0}^{\infty} \alpha^{2^n}$$

is transcendental.

Proof: Consider $f(x), (f(x))^2, (f(x))^3, \ldots$, which are power series with rational integer coefficients. So, we can write

$$(f(x))^k = \sum_{n=0}^{\infty} b_{k,n} x^n$$

with $b_{k,n} \in \mathbb{Z}$. Let $m \in \mathbb{N} \cup \{0\}$ be fixed. We will choose a particular value for it later. Now, there exist polynomials

$$P_i(x) := \sum_{n=0}^{m} a_{i,n} x^n \in \mathbb{Z}[x], \quad i = 0, 1, \ldots, m$$

which are not all zero. There also exists a power series $g_m(x)$ such that

$$P_m(x)(f(x))^m + P_{m-1}(x)(f(x))^{m-1} + \cdots + P_1(x)f(x) + P_0(x) = x^{m^2} g_m(x). \tag{$*1$}$$

Now, we have that

$$\sum_{i=0}^{m} \sum_{k=0}^{\min\{m,n\}} a_{i,k} b_{i,n-k} = 0 \tag{$*2$}$$

for $n = 0, 1, \ldots, m^2 - 1$. This gives us a set of equations with $(m+1)^2$ unknown values, which is more than the m^2 solutions. Hence there is a nonzero solution in $\mathbb{Q}$. We can multiply through the solution to clear denominators and get a solution in $\mathbb{Z}$.

Now we claim that g_m is not identically zero. To see this, suppose toward a contradiction that it is. Now, since the P_i are not all zero, we have that f would be algebraic, which contradictions **Lemma 1**. Hence, we have that

$$g_m(x) = x^\sigma h_m(x) \tag{$*3$}$$

with $\sigma \geq 0$ and $h_m(0) \neq 0$.

Suppose toward a contradiction that $f(\alpha)$ is algebraic. Let $K = Q(\alpha, f(\alpha))$ and denote $d := \big[K : \mathbb{Q}\big]$. Let $a := \text{den}(\alpha)$ be the denominator of α, such that

$$\alpha = \frac{\beta}{a}$$

where β is an algebraic integer of K. Let $b := \text{den}(f(\alpha))$, such that

$$f(\alpha) = \frac{r}{b}$$

where r is an algebraic integer of K as well.

For integers $n \geq 1$, we have that

$$f\left(x^{2^n}\right) = f(x) - \sum_{k=0}^{n-1} x^{2^k}. \tag{$*4$}$$

For all $n \in \mathbb{N} \cup \{0\}$, we see that

$$f\left(\alpha^{2^n}\right) = \frac{A_n}{b a^{2^{n-1}}},$$

where A_n is an algebraic integer of K.

Replacing x by α^{2^n} in $(*1)$ and $(*3)$, we have that

$$\sum_{i=0}^{m} P_i\left(\frac{\beta^{2^n}}{a^{2^n}}\right)\left(\frac{A_n}{b a^{2^{n-1}}}\right)^i = \alpha^{(m^2+\sigma)2^n} h_m\left(\alpha^{2^n}\right). \tag{$*5$}$$

Denote B_n to be the left-hand side of $(*5)$. Observe that $B_n \in K$.

Since the P_i have integer coefficients and since $\deg(P_i) \leq m$, we have that

$$\text{den}(B_n) \leq a^{m2^n} b^m a^{m2^{n-1}} \leq b^m a^{m2^{n+1}}. \tag{$*6$}$$

Now let $n \to \infty$ in $(*5)$. Since $h_m(0) \neq 0$, we have that

$$B_n \sim h_m(0) \alpha^{(m^2+\sigma)2^n}.$$

So, for very large n, we have that $B_n \neq 0$. Additionally, there exists $c_1 > 0$ independent of n such that for any integer $n \in \mathbb{N}$, we have

$$|B_n| \leq c_1 |\alpha|^{m^2 2^n}. \tag{$*7$}$$

By **Theorem 2**, we get

$$\overline{|B_n|} \leq \sum_{i=0}^{m} \left(\left(\sum_{j=0}^{m} \left(\overline{|a_{i,j}|}\, \overline{|\alpha|}^{j2^n} \right) \right) \left(\overline{|f\left(\alpha^{2^n}\right)|} \right)^i \right).$$

Let $c_2 := \max\limits_{i,j} |a_{i,j}|$ and $c_3 := \max\{2, \overline{|\alpha|}, \overline{|f(\alpha)|}\}$. Then,

$$\overline{|B_n|} \leq c_2 (m+1)^2 c_3^{m2^{n+1}}.$$

(To see this, use ($*4$) and **Theorem 2**.)

Since $B_n \neq 0$, we can use size inequality, and since $B_n \in K$, we have $\deg(B_n) \leq d$. Using ($*6$) and ($*7$), we have

$$c_1 |\alpha|^{m^2 2^n} \geq \left(c_2 (m+1)^2 c_3^{m2^{n+1}} \right)^{-d+1} \left(b^m a^{m2^{n+1}} \right)^{-d}.$$

Next we take the logarithm of both sides. This yields

$$-md\log(b) + (-d+1)\log(c_2(m+1)^2) + 2^n\big(-m^2 \log(\alpha) + 2m(-d+1)\log(c_3) - d\log(a)\big) \leq \log(c_1). \tag{$*8$}$$

Since $|\alpha| < 1$, we can choose the value of m (which we said we would choose later) such that

$$-m^2 \log(|\alpha|) + 2m\big((-d+1)\log(c_3) - d\log(a)\big) > 0.$$

We can see that as $n \to \infty$, the left-hand side of ($*8$) also goes to infinity, which is a contradiction because it has an upper bound. Therefore, the assumption that $f(\alpha)$ is algebraic is false. This proves the theorem. $\square$

Appendix G

Ali Uncu: Ramanujan Sums

First we define the Mobius function μ.

Definition: Let $n = p_1^{a_1} \cdots p_r^{a_r}$. Then, define

$$\mu(n) := \begin{cases} (-1)^{a_1+\cdots+a_r}, & a_1 = \cdots = a_r 1 \\ 0, & \text{otherwise} \end{cases} .$$

Example: $\mu(2) = -1$, $\mu(4) = 0$, and $\mu(6) = 1$.

Definition: An arithmetic function f is said to be periodic with period k if

$$f(n+k) = f(n)$$

for all $n \in \mathbb{N}$. The smallest such k is called the fundamental period.

Theorem: For some fixed $k \geq 1$, we have

$$g(n) = \sum_{m=0}^{k-1} e^{2\pi i m n/k} = \begin{cases} 0, & k \nmid n \\ k, & k \mid n \end{cases} .$$

Proof: Observe that

$$f(x) = \sum_{m=0}^{k-1} x^m = \begin{cases} \dfrac{x^k - 1}{x-1}, & x \neq 1 \\ k, & x = 1 \end{cases} . \ \square$$

Remark: We want to show that we can identify every periodic arithmetic function as a Fourier series.

Theorem: (Lagrange Interpolation Theorem) Let $z_0, \ldots, z_{k-1}$ be all distinct complex numbers and let $w_0, \ldots, w_{k-1} \in \mathbb{C}$. Then, there exists a unique polynomial $P(x)$ of degree $\leq k-1$ such that $P(z_m) = w_m$ for $m = 0, 1, \ldots, k-1$.

Proof: Consider

$$A_m := \prod_{\substack{n=0 \\ m \neq n}}^{k-1} (z - z_n)$$

and

$$P(z) := \sum_{m=0}^{k-1} w_m \frac{A_m(z)}{A_m(z_m)}.$$

Thus existence is shown. Assume that $Q(z)$ is a polynomial of degree $\leq k-1$ such that $Q(z_m) = w_m$ for all m. Then, $P(z)-Q(z) = 0$ at k points, and this implies that $P = Q$, hence uniqueness is shown. $\square$

Remark: Now, we apply this, picking $z_m := e^{2\pi im/k}$.

Theorem: Given k complex numbers $w_0, \ldots, w_{k-1}$, there exists k uniquely determined $a_0, \ldots, a_{k-1} \in \mathbb{C}$ such that

$$w_m = \sum_{m=0}^{k-1} a_m e^{2\pi imn/k}.$$

Moreover,

$$a_m = \frac{1}{k} \sum_{m=0}^{k-1} w_m e^{-2\pi imn/k}.$$

Proof: The previous theorem gives us the first conclusion. Now, take the expressions for each w_r, multiply each by $e^{-2\pi imr/k}$, and sum over m. This gives us

$$\sum_{m=0}^{k-1} w_m e^{-2\pi imr/k} = \sum_{n=0}^{k-1} a_n \underbrace{\sum_{m=0}^{k-1} e^{2\pi im(n-r)/k}}_{=\begin{cases} 0, & k \nmid n-r \\ k, & k \mid n-r \end{cases}}.$$

We also know $|n-r| \leq k-1$. So,

$$\sum_{m=0}^{k-1} w_m e^{-2\pi imr/k} = ka_r. \ \square$$

Corollary: Let f be an arithmetic function which is periodic with period k. Then, there exists a unique arithmetic function g which is periodic with period k such that

$$f(n) = \sum_{m \bmod k} g(m) e^{2\pi imn/k}$$

and

$$g(m) = \frac{1}{k} \sum_{n \bmod k} f(n) e^{-2\pi imn/k}.$$

Remark: We want to understand the Ramanujan Sum

$$c_k(n) = \sum_{\substack{m \bmod k \\ \gcd(m,k)=1}} e^{2\pi imn/k}.$$

Observe that $c_k(1) = \mu(k)$ and $c_k(k) = \varphi(k)$. We will prove that

$$c_k(n) = \sum_{d \mid \gcd(n,k)} d\mu\left(\frac{k}{d}\right).$$

Theorem: Let

$$s_k(n) := \sum_{d \mid \gcd(n,k)} f(d) g\left(\frac{k}{d}\right).$$

Then, $s_k(n)$ is periodic with period k. So,

$$s_k(n) = \sum_{m \bmod k} a_k(m) e^{2\pi imn/k}$$

where

$$a_k(m) = \sum_{d|\gcd(m,k)} g(d) f\left(\frac{k}{d}\right) \cdot \frac{d}{k}.$$

Proof: Consider (using $n = dc$)

$$\begin{aligned}
a_k(m) &= \frac{1}{k} \sum_{n \bmod k} s_k(n) e^{-2\pi imn/k} \\
&= \frac{1}{k} \sum_{n=1}^{k} \sum_{d|\gcd(n,k)} f(d) g\left(\frac{d}{k}\right) e^{-2\pi imn/k} \\
&= \frac{1}{k} \sum_{d|k} f(d) g\left(\frac{k}{d}\right) \sum_{c=1}^{k/d} e^{-2\pi imdc/k} \\
&= \frac{1}{k} \sum_{d|k} f\left(\frac{k}{d}\right) g(d) \sum_{c=1}^{d} e^{-2\pi imc/d}, \text{ when } d \mid m, \\
&= \frac{1}{k} \sum_{d|k} f\left(\frac{k}{d}\right) g(d) d \text{ when } d \mid m. \ \square
\end{aligned}$$

Corollary: $c_k(n) = \sum_{d|\gcd(n,k)} d\mu\left(\frac{k}{d}\right)$

Proof: Define $f(n) = n$ and $g(k) = \mu(k)$. Then,

$$\sum_{d|\gcd(n,k)} d\mu\left(\frac{k}{d}\right) = \sum_{m=1}^{k} a_k(m) e^{2\pi imn/k}$$

where

$$a_k(m) = \sum_{d|\gcd(m,k)} \mu(d) = \left\lfloor \frac{1}{\gcd(n,k)} \right\rfloor = \begin{cases} 1, & \gcd(n,k) = 1 \\ 0, & \gcd(n,k) > 1 \end{cases}.$$

Observe that

$$\sum_{d|n} \mu(d) = \left\lfloor \frac{1}{n} \right\rfloor.$$

So,

$$\sum_{d|\gcd(n,k)} d\mu\left(\frac{k}{d}\right) = \sum_{\substack{m \bmod k \\ \gcd(m,k)=1}} e^{2\pi imn/k} = c_k(m). \ \square$$

Theorem: Define

$$s_k(n) := \sum_{d|\gcd(n,k)} f(d) g\left(\frac{k}{d}\right).$$

If f and g are both multiplicative, then

(i) $s_{mk}(ab) = s_m(a) s_k(b)$ whenever $\gcd(a,k) = \gcd(b,m) = 1$.

(ii) $s_m(ab) = s_m(a)$ if $\gcd(b, m) = 1$.

(iii) $s_{mk}(a) = s_m(a)$??? if $\gcd(n, k) = 1$.

Proof: For (i), let $\gcd(a, k) = \gcd(b, m) = 1$. Then, we can say that

$$\gcd(ab, mk) = \gcd(a, m)\gcd(b, k).$$

Now, we can write (with $d = d_1 d_2$)

$$\begin{aligned}
s_{mk}(ab) &= \sum_{d|\gcd(ab,mk)} f(d)g\left(\frac{k}{d}\right) \\
&= \sum_{d|\gcd(a,m)\gcd(b,k)} f(d)g\left(\frac{k}{d}\right) \\
&= \sum_{\substack{d_1|\gcd(a,m) \\ d_2|\gcd(b,k)}} f(d_1 d_2)g\left(\frac{mk}{d_1 d_2}\right) \\
&= \left(\sum_{d_1|\gcd(a,m)} f(d_1)g\left(\frac{m}{d_1}\right)\right)\left(\sum_{d_2|\gcd(b,k)} f(d_2)g\left(\frac{k}{d_2}\right)\right) \\
&= s_m(a)s_k(b).
\end{aligned}$$

To see (ii), set $k = 1$ in (i) and the result follows. To see (iii), set $b = 1$ in (ii).

Appendix H

Hongyan Hou: Bernoulli Coefficients

Recall the sums

$$\sum_{n=1}^{N} n = \frac{N(N+1)}{2},$$

$$\sum_{n=1}^{N} n^2 = \frac{N(N+1)(2N+1)}{6},$$

$$\sum_{n=1}^{N} n^3 = \frac{N^2(N+1)^2}{4}.$$

This was generalized by Jacob Bernoulli as

$$\binom{x}{j} = \frac{x(x-1)\cdots(x-j+1)}{j!}.$$

Now, we define $A_{i,j}$ by

$$x^q = \sum_{j\geq 0} A_{q,j}\binom{x}{j} \tag{1}$$

for all q. We have that $A_{0,0} = 1$ and $A_{0,j} = 0$ for all $j > 0$.

Set $x = 0$ and consider $q \geq 1$. Then,

$$0 = A_{q,0} + \sum_{j\geq 1} A_{q,j}\binom{0}{j}$$

and so $A_{q,0} = 0$ for all q.

Observe that

$$A_{q,q}\binom{x}{q} = A_{q,q}\frac{x(x-1)\cdots(x-q+1)}{q!}$$

and so $A_{q,q} = q!$.

Now we find a general expression of $A_{q,j}$. In (1), let $x = \ell = 0, 1, 2, \ldots, k \leq q$. Multiply both sides by

$$(-1)^{\ell}\binom{k}{\ell}.$$

Taking sums,

$$\begin{aligned}
\sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\ell^{q} &= \sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\sum_{j=0}^{q}A_{q,j}\binom{\ell}{j}\\
&= \sum_{j=0}^{q}A_{q,j}\sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\binom{\ell}{j}\\
&= \sum_{j=0}^{q}A_{q,j}\sum_{j\le\ell\le k}(-1)^{\ell}\binom{k}{\ell}\binom{\ell}{j}\\
&= \sum_{j=0}^{q}A_{q,j}\sum_{j\le\ell\le k}(-1)^{\ell}\frac{k!}{\ell!(k-\ell)!}\frac{\ell!}{j!(\ell-j)!}\\
&= \sum_{j=0}^{q}A_{q,j}\frac{k!}{j!(k-j)!}\sum_{j\le\ell\le k}(-1)^{\ell}\frac{(k-j)!}{(k-\ell)!(\ell-j)!}\\
&= \sum_{j=0}^{q}A_{q,j}\binom{k}{j}\sum_{j\le\ell\le k}(-1)^{\ell}\binom{k-j}{\ell-j}.
\end{aligned}$$

Now, if $j < k$¡ then making the substitution $\lambda = \ell - j$, we have

$$\begin{aligned}
\sum_{j\le k\le\ell}(-1)^{\ell}\binom{k-j}{\ell-j} &= \sum_{\lambda=0}^{k-j}(-1)^{\lambda+j}\binom{k-j}{\lambda}\\
&= (-1)^{j}\sum_{\lambda=0}^{k-j}(-1)^{\lambda}\binom{k-j}{\lambda}\\
&= (-1)^{j}(1-1)^{k-j}\\
&= 0.
\end{aligned}$$

If $j = k$, then

$$\sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\ell^{q} = (-1)^{k}A_{q,k}.$$

Bernoulli Polynomials

Define

$$B_{\gamma}(y) := \gamma\sum_{j\ge0}A_{\gamma-1,+j}\binom{y}{j+1} + c_{\gamma}$$

for $\gamma = 1, 2, \ldots$. Now,

$$B_{\gamma}(0) = \gamma\sum_{j\ge0}A_{\gamma-1,j}\binom{0}{j+1} + c_{\gamma} = c_{\gamma}.$$

Also,

$$B_{\gamma}(1) = \gamma\sum_{j\ge0}A_{\gamma-1,j}\binom{1}{j+1} + c_{\gamma} = \gamma A_{\gamma-1,0} + c_{\gamma} = c_{\gamma}$$

for $\gamma > 1$.

Clearly,

$$B_{1}(1) = \sum_{j\ge0}A_{0,j}\binom{1}{j+1} + c_{1} = A_{0,0} + c_{1} = 1 + c_{1}.$$

Now, recall the identity

$$\binom{x}{j} = \binom{x+1}{j+1} - \binom{x}{j+1}$$

and so

$$\begin{aligned} x^q &= \sum_{j\geq 0} A_{q,j}\binom{x}{j} \\ &= \sum_{j\geq 0} A_{q,j}\left[\binom{x+1}{j+1} - \binom{x}{j+1}\right] \\ &= \frac{1}{q+1}[B_{q+1}(x+1) - B_{q+1}(x)]. \end{aligned}$$

Differentiating both sides yields

$$qx^{q-1} = \frac{1}{q+1}[B'_{q+1}(x+1) - B'_{q+1}(x)].$$

Observe that

$$\frac{1}{q+1}B'_{q+1}(x+1) - B_q(x+1) = \frac{1}{q+1}B'_{q+1}(x) - B_q(x)$$

and so

$$\frac{1}{q+1}B'_{q+1}(x) - B_q(x) = k_q$$

with period 1. Thus

$$\frac{1}{q+1}B'_{q+1}(x) = B_q(x).$$

Next,

$$\begin{aligned} B_1(x) &= \sum_{j\geq 0} A_{0,j}\binom{x}{j+1} + c_1 \\ &= A_{0,0}\binom{x}{1} + c_1 \\ &= x + c_1. \end{aligned}$$

Hence by the earlier formula,

$$\frac{1}{2}B'_2(x) = x + c_1$$

so with the initial condition $B_2(0) = c_2$, we get

$$\frac{1}{2}B_2(x) = \frac{x^2}{2} + \frac{c_1}{1!}x + \frac{c_2}{2!}.$$

Iterating this process,

$$\frac{1}{q!}B_q(x) = \frac{x^q}{q!} + \frac{c_1}{1!}\cdot\frac{x^{q-1}}{(q-1)!} + \cdots + \frac{c_{q-1}}{(q-1)!}\cdot\frac{x}{1!}.$$

We find a recursion formula for $c_q =: B_q$ which we define to be the Bernoulli numbers. The first few are

$$B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_3 = 0.$$

Appendix I

Frank Patane: Bernoulli Polynomials

Recall: Recall that we define Bernoulli Polynomials by

$$B_q(x) := \sum_{k=0}^{q} \binom{q}{k} B_k(0) x^{q-k} = (B+x)^q$$

where we write B_q as B^q.

Remark: Observe that

$$B_1(x) = x - \frac{1}{2} \implies B_1 = -\frac{1}{2}, B_2(x) = x(x-1) + \frac{1}{6} \implies B_2 = \frac{1}{6}, B_3(x) = x(x-1)\left(x - \frac{1}{2}\right) \implies B_3 = 0, B_4(x)$$

So, we see that $B_k := B_k(0) = B_k(1)$ if $k > 1$.

Remark: We have that

$$\sum_{n=1}^{N-1} n^q = \frac{1}{q+1}\left[B_{q+1}(N) - B_{q+1}(1)\right], \quad q \geq 1.$$

So, for $q = 2$:

$$\sum_{n=1}^{N-1} n^2 = \frac{1}{3} B_3(N) = \frac{1}{3} \cdot N(N-1)\left(N - \frac{1}{2}\right).$$

For $q = 3$:

$$\sum_{n=1}^{N-1} n^3 = \frac{1}{4} \cdot N^2(N-1)^2.$$

Also,

$$\sum_{n=1}^{N-1} n^3 = \left(\sum_{n=1}^{N-1} n\right)^2$$

, and we see that for $q = 1$:

$$\sum_{n=1}^{N-1} n = \frac{1}{2} \cdot N(N-1).$$

Claim: We prove now that

$$\left(\sum_{n=1}^{N} n\right)^2 = \sum_{n=1}^{N} n^3.$$

Well,

$$\begin{aligned} 1^3 + 2^3 + 3^3 + \cdots + N^3 &= 1 + \underbrace{(3+5)}_{2^3} + \underbrace{(7+9+11)}_{3^3} + \cdots \\ &= (\text{number of terms})^2 \\ &= \left(\sum_{n=1}^{N} n \right)^2 . \end{aligned}$$

Remark: We have shown that $B_q(1-x) = (-1)^q B_q(x)$, so for $x = 0$, we see that

$$B_{2k+1}(1) = -B_{2k+1}(0).$$

The Zeros of $B_q(x)$

We have shown that $B_{2k+1}(x)$ for $k \geq 1$ has only the zeros 0, $\frac{1}{2}$, 1 in $[0,1]$. By the above remark

$$B_{2k+1}\left(\frac{1}{2}\right) = -B_{2k+1}\left(\frac{1}{2}\right).$$

Suppose that $B_{2k+1}(x)$ has 4 distinct zeros $0, \alpha_1, \alpha_2, 1$ in $[0,1]$. Then by **Rolle's Theorem**,

$$B'_{2k+1}(x) = (2k+1)B_{2k}(x)$$

has zeros $\beta_1, \beta_2, \beta_3$, with

$$0 < \beta_1 < \alpha_1 < \beta_2 < \alpha_2 < \beta_3 < 1.$$

If we repeat the process, $B_{2k-1}(x)$ will have two zeros γ_1, γ_2, with

$$\beta_1 < \gamma_1 < \beta_2 < \gamma_2 < \beta_3$$

so it has zeros in $[0,1]$. But, $B_3(x)$ has degree three, and so this is a contradiction. Therefore, $0, \frac{1}{2}, 1$ are all the zeros of $B_{2k+1}(x)$ for $k \geq 1$.

Corollary: $B_{2k}(x) - B_{2k}(0)$ does not change sign in $(0,1)$.

Proof: Let α be a zero in $(0,1)$ of $B_{2k}(x) - B_{2k}(0)$. Together with the fact that

$$B_{2k}(0) - B_{2k}(0) = B_{2k}(1) - B_{2k}(0) = 0$$

this implies that $B_{2k-1}(x)$ would have four zeros in $[0,1]$, which is a contradiction.

Now we consider the case $2k+1$. Note that $B_{2k+1}(0) = B_{2k+1}(\frac{1}{2}) = 0$, and so B_{2k} has a zero β in $(0, \frac{1}{2})$. Thus

$$(B_{2k}(x) - B_{2k}(0))(B_{2k}(\beta) - B_{2k}) > 0$$

and so

$$\left(\binom{2k}{2} B_{2k-2}(0)x^2 + \binom{2k}{4} B_{2k-4}(0)x^4 + \cdots \right) \cdot B_{2k}(0) < 0.$$

Therefore, $B_{2k-2}(0) \cdot B_{2k}(0) < 0$. □

Remark: Also note that for $k > 2$, the quantity $B_{2k}(x) - B_{2k}(0)$ vanishes at 0 and 1, yet $0, 1$ are also roots of $B_{2k-1}(x)$ (and so are double roots). In particular, $B_2(x) - B_2(0) = x(x-1)$. Thus, $(B_2(x) - B_2(0))^2$ is a factor of $B_{2k}(x) - B_{2k}(0)$ for $k \geq 2$. Thus, $B_4(x) - B_4(0) = (B_2(x) - B_2(0))^2$.

Fourier Expansion of Bernoulli Polynomials

Definition: Define

$$\psi_q(t) := B_q(t - \lfloor t \rfloor),$$

with $q \geq 1$ has period 1. Also, $\psi_q(t)$ for $q \geq 2$ is continuous since $B_q(0) = B_1(q)$. Then, we have that

$$\psi_q(t) = \frac{a_0^{(q)}}{2} + \sum_{n=1}^{\infty} a_n^{(q)} \cos(2\pi n t) + b_n^{(q)} \sin(2\pi n t),$$

where

$$a_n^{(q)} := 2 \int_0^1 B_q(x) \cos(2\pi n x)\, dx$$

and

$$b_n^{(q)} := 2 \int_0^1 B_q(x) \sin(2\pi n x)\, dx.$$

Consider: First consider the case $n = 0$. Then,

$$a_0^{(q)} = 2 \int_0^1 B_q(x)\, dx = \frac{2}{q+1} \int_0^1 B'_{q+1}(x)\, dx = 0.$$

For $n > 0$, we can integrate by parts to get

$$\begin{aligned} a_n^{(q)} &= 2 \left[\left[\underbrace{\frac{B_q(x) \sin(2\pi n x)}{2\pi n}}_{=0} \Big|_0^1 \right] - \frac{1}{2\pi n} \int_0^1 B'_q(x) \sin(2\pi n x)\, dx \right] \\ &= -\frac{2q}{2\pi n} \int_0^1 B_{q-1} \sin(2\pi n x)\, dx \\ &= \frac{-q}{2\pi n} b_n^{(q-1)}. \end{aligned}$$

Thus, $a_n^{(1)} = 0$ for $n \geq 1$, since

$$b_n^{(0)} = 2 \int_0^1 \sin(2\pi n x)\, dx = 0.$$

Similarly by integration by parts, we get that

$$b_n^{(q)} = \frac{q}{2\pi n} q_n^{(q-1)}, \quad n \geq 1.$$

Lastly,

$$b_n^{(1)} = \frac{-2}{2\pi n}.$$

Thus for $k \geq 1$ and $n > 0$, we get that

$$a_n^{(2k-1)} = 0,$$

$$b_n^{(2k)} = 0,$$

$$a_n^{(2k)} = (-1)^{k-1} \frac{2(2k)!}{(2\pi n)^{2k}},$$

$$b_n^{(2k-1)} = (-1)^k \frac{2(2k-1)!}{(2\pi n)^{2k-1}}.$$

Remark: Concluding,

$$\begin{aligned}\psi_{2k-1}(t) &= \sum_{n=1}^{\infty}(-1)^k \frac{2(2k-1)!}{(2\pi n)^{2k-1}} \cdot \sin(2\pi nt) \\ &= 2(-1)^k(2k-1)! \sum_{n=1}^{\infty} \frac{\sin(2\pi nt)}{(2\pi n)^{2k-1}}.\end{aligned}$$

Similarly,

$$\psi_{2k}(t) = 2(-1)^{k-1}(2k)! \sum_{n=1}^{\infty} \frac{\cos(2\pi nt)}{(2\pi n)^{2k}}.$$

Appendix J

Todd Molnar: The Sathe/Selberg-Delange Theorem

Lemma 1: If

$$D(s) := \sum_{n=1}^{\infty} \frac{d_n}{n^s}$$

is absolutely convergent for $\Re(s) = \sigma > \sigma_a$ and $c > \sigma_a$, then for $x > 0$,

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} D(s) \cdot \frac{x^s}{s}\, ds = \begin{cases} \sum_{n \geq x} d_n, & x \notin \mathbb{Z} \\ \frac{d_x}{2} + \sum_{n<x} d_n, & x \in \mathbb{Z} \end{cases}.$$

Theorem 2: (Hankel's Formula) Let H be the contour which travels along $y = -1$ from $x = -r$ toward the y-axis, then when it intersects a circle centered at zero, it follows it until it is on the line $x = r$ (on the left-hand side of the y-axis), then it follows this horizontal line back to $-\infty$. Then, for all $z \in \mathbb{C}$, we have that

$$\frac{1}{\Gamma(z)} = \frac{1}{2\pi z} \int_H s^{-z} e^s\, ds$$

where $\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1}\, dt$.

Theorem 3: (Cauchy's Estimate) Let f be analytic in a circle centered at $a \in \mathbb{C}$ of radius R, and assume $|f(z)| \leq M$ in this circle; then:

$$|f^{(n)}(a)| \leq \frac{n! \cdot M}{R^n}.$$

Theorem 4: For all $s \in \mathbb{C}$ such that $\Re(s) =: \omega > 0$, we have that $\zeta(s)$ is a meromorphic function, with a simple pole at $s = 1$ with residue 1. Furthermore, there exists $c > 0$ such that $\zeta(s) \neq 0$ in the region:

$$\omega \geq 1 - \frac{c}{1 + \log|t|}. \tag{1}$$

Definition: Let $Z(s; z) := (s-1)^z \zeta^z(s) \cdot s^{-1}$, and choose the principal branch of the complex logarithm, such that $Z(1; z) = 1$.

Lemma 2: $Z(s; z)$ is holomorphic in the disk $|s - 1| < 1$ and

$$Z(s; z) = \sum_{j=0}^{\infty} \gamma_j(z) \cdot \frac{(s-1)^j}{j!}, \tag{2}$$

where the $\gamma_j(z)$ are entire functions of $z \in \mathbb{C}$, such that for all $A > 0$ and $\epsilon > 0$:

$$\frac{\gamma_j(z)}{j!} <<_{A,\epsilon} (1+\epsilon)^k, \quad |z| \leq A. \tag{3}$$

Definition: Let D denote the simply connected domain obtained by deleting the real segment $[1-c, 1]$ from the region in (1), admitting the analytic continuation:

$$\zeta^z(s) = \frac{s \cdot Z(s;z)}{(s-1)^z}, \quad s \in D.$$

Remark: For all $A > 0$, we have that

$$|\zeta^z(s)| <<_A (1+\log|t|)^A, \text{ for } |z| \leq A,\ s \in D, \text{ and } |s-1| >> 1. \tag{4}$$

Definition: Let $z \in \mathbb{C}$, $c_0 > 0$, $0 < \delta \leq 1$, $M > 0$. We sat that

$$F(s) := \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

is of <u>type</u> $P(z, c_0, \delta, M)$ if $G(s;z) = F(s)\zeta^z(s)$ may be continued to a holomorphic function for $\sigma \geq 1 - \frac{c_0}{1+\log|t|}$ and

$$|G(s;z)| \leq M(1+|t|)^{1-\delta} \tag{5}$$

in this domain.

Definition: If $F(s)$ is of type $P(z, c_0, \delta, M)$ and there exists a sequence of positive real numbers $\{b_n\}_{n\in\mathbb{N}}$ such that $|a_n| \leq b_n$ and if $\sum_{n=1}^{\infty} \frac{b_n}{n^s}$ is of type $P(w, c_0, \delta, M)$, some $w \in \mathbb{C}$, we say that $F(s)$ is of <u>type</u> $T(z, w, c_0, \delta, M)$.

Definition: In the domain where $G(s;z)$ is holomorphic, set

$$G^{(k)}(s;z) := \frac{\partial^k}{\partial s^k} G(s;z)$$

and

$$\lambda_k(z) := \frac{1}{\Gamma(z-k)} \cdot \sum_{h+j=k} \frac{1}{h! \cdot j!} G^{(k)}(1;z) \cdot \gamma_j(z).$$

Theorem 5: Let $F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$ be a Dirichlet series of type $T(z, w, c_0, \delta, M)$. For $x \geq 3$, $N \geq 0$, $A > 0$, $|z| \leq A$, and $|w| \leq A$, we have

$$\sum_{n \leq x} a_n = x \cdot \log^{z-1}(x) \cdot \left[\sum_{k=0}^{N} \frac{\lambda_k(z)}{\log^k(x)} + O\left(M \cdot R_n(x)\right) \right]$$

where

$$R_n(x) = e^{-c_1\sqrt{\log(x)}} + \left(\frac{c_2 N + 1}{\log(x)}\right)^{N+1}.$$

The constants $c_1, c_2 > 0$ and the implicit constant in the Landau Symbol depend at most on c_0, δ, A.

Definition: Define

$$\omega(n) := \sum_{p|n} 1.$$

Observe that if $\gcd(a,b) = 1$, then $\omega(ab) = \omega(a) + \omega(b)$. Therefore,

$$z^{\omega(ab)} = z^{\omega(a)} z^{\omega(b)}.$$

Define and calculate

$$\begin{aligned} F_1(s;z) &:= \sum_{n=1}^{\infty} \frac{z^{\omega(n)}}{n^2} \\ &= \prod_{p \text{ prime}} \left(1 + \frac{z}{p^s} + \frac{z}{p^{2s}} + \frac{z}{p^{3s}} + \cdots\right) \\ &= \prod_{p \text{ prime}} \left[1 + z\left(\frac{1}{1-p^{-s}} - 1\right)\right] \\ &= \prod_{p \text{ prime}} \left[1 + z\left(\frac{p^{-s}}{1-p^{-s}}\right)\right] \\ &= \prod_{p \text{ prime}} \left[1 + \frac{z}{p^s - 1}\right]. \end{aligned}$$

Now, using the following identity for the Riemann Zeta Function

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}$$

we have that

$$\begin{aligned} G_1(s;z) &= F_1(s;z)\zeta^{-z}(s) \\ &= \prod_{p \text{ prime}} \left(1 + \frac{z}{p^s - 1}\right)\left(1 - \frac{1}{p^s}\right)^z \\ &= \sum_{n=1}^{\infty} \frac{b_{1,z}(n)}{n^s}. \end{aligned}$$

Note that $b_{1,z}(n)$ is multiplicative.

Next,

$$1 + \sum_{v=1}^{\infty} b_{1,z}(p^\nu)\xi^\nu = \left(1 + \frac{\xi z}{1-\xi}\right)(1-\xi)^z, \quad |\xi| < 1.$$

In particular,

$$b_{1,z}(p) = 0. \tag{1*}$$

Cauchy's Estimate for $|z| \le A$ gives

$$|b_{1,z}(p^\nu)| \le M \cdot 2^{\nu/2}, \quad \nu \ge 2. \tag{2*}$$

Where

$$M := \sup_{\substack{|z| \le A \\ |\xi| \le \frac{1}{\sqrt{2}}}} \left|\left(1 + \frac{\xi z}{1-\xi}\right)(1-\xi)^z\right|.$$

From (1∗) and (2∗), for $\sigma > \frac{1}{2}$, we get that

$$\sum_{p \text{ prime}} \sum_{\nu=1}^{\infty} \frac{|b_{1,z}(p^\nu)|}{p^{\nu\sigma}} \leq 2 \cdot M \sum_{p \text{ prime}} \frac{1}{p^\sigma(p^\sigma - \sqrt{2})} \leq \frac{c \cdot M}{\sigma - \frac{1}{2}},$$

where $c > 0$ is a constant.

Therefore, $G_1(s; z)$ is absolutely convergent for $\sigma > \frac{1}{2}$, and in the case $\sigma > \frac{3}{4}$, it follows that

$$G_1(s; z) <<_A 1.$$

This completes the theorem. □

Corollary: If $\theta \in \mathbb{R} \setminus \mathbb{Q}$, then $\{\omega(n) \cdot \theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Proof: For all $A > 0$, there exists $c_1 = c_1(A)$ and $c_2 = c_2(A)$ such that uniformly for $x \geq 3$, $N \geq 0$, and $|z| \leq A$, we have

$$\sum_{n\leq x} z^{\omega(n)} = x \log^{z-1}(x) \left[\sum_{k=0}^{N} \frac{\lambda_k(z)}{\log^k(x)} + O_A\left(e^{-c_1\sqrt{\log(x)}} + \left(\frac{c_2 N + 1}{\log(x)}\right)^{N+1}\right)\right]. \tag{3∗}$$

From (3∗) it follows that using $z = e^{2\pi i h\theta}$ and $\theta \in \mathbb{R} \setminus \mathbb{Q}$ and $h \neq 0$, we get $|z| = 1$ and $z \neq 1$. Hence,

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n\leq x} z^{\omega(n)}.$$

By **Weyl's Criterion**, this shows that $\{\omega(n)\theta\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1. □

Proof of Theorem 5: Let $0 < c < c_0$ such that $\zeta(s) \neq 0$ in $\sigma \geq 1 - \frac{c}{1 + \log|t|}$ by **Theorem 4**. Then, by (4) and (5), we have that

$$F(s) <<_A M(1 + \log|t|)^A(1 + |t|)^{1-\delta} <<_{A,\delta} M(1 + |t|)^{1-\delta/2} \tag{6}$$

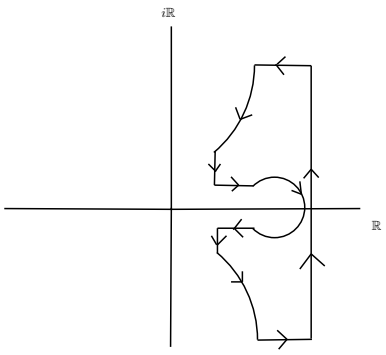
uniformly for $s \in S$ with $|s - 1| >> 1$ and $|z| \leq A$. Setting

$$A(x) := \sum_{n\leq x} a_n$$

we apply **Lemma 1** to see that

$$\int_0^x A(t)\, dt = \frac{1}{2\pi i} \int_{K-i\infty}^{K+i\infty} F(s) \cdot \frac{x^{s+1}}{s(s+1)}\, ds,$$

with $K = 1 + \frac{1}{\log(x)}$. Let $T > 1$. Consider the contour integral given by the residue theorem:



The circle about $s = 1$ has radius $r = \dfrac{1}{2\log(x)}$ and $\sigma(T) = 1 - \dfrac{c}{2(1+\log|t|)}$. By (6), contributions from $[K \pm iT, K \pm i\infty]$ are $<<_{A,\delta} M \cdot x^2 \cdot T^{-\delta/2}$, and the same holds true for $[\sigma(T) \pm iT, K \pm iT]$.

With $\sigma = \sigma(T)$ and $T = \exp\left(\sqrt{\dfrac{c}{\delta}\log(x)}\right)$ and $x \geq x_0$, it follows that

$$\int_0^x A(t)\, dt = \Phi(x) + O(Mx^2 e^{-c_3\sqrt{\log(x)}}) \tag{7}$$

and so

$$\Phi(x) = \frac{1}{2\pi i}\int_{\mathscr{C}} F(s) \cdot \frac{x^{s+1}}{s(s+1)}\, ds.$$

Thus,

$$\Phi'(x) = \frac{1}{2\pi i}\int_{\mathscr{C}} F(s) \cdot \frac{x^s}{s}\, ds,$$

and

$$\Phi''(x) = \frac{1}{2\pi i}\int_{\mathscr{C}} F(s) \cdot x^{s-1}\, dx.$$

The result follows from some further calculation which is not included in this talk. $\square$

Index