

Contents

1	Introduction	3
1.1	The pioneers in this field	4
1.2	Some basic ideas	5
1.3	Background Information	8
1.4	Abstract	9
1.4.1	Chapter 2	9
1.4.2	Chapter 3	9
1.4.3	Chapter 4	10
2	Approximation of real and algebraic numbers	11
2.1	Approximation of real numbers by algebraic numbers	11
2.2	Simultaneous Approximation	20
2.3	Approximation of algebraic numbers by rational numbers	26
2.4	Approximation of algebraic numbers by algebraic numbers	32
2.5	Further refinements and generalizations of Liouville's Theorem	43
3	Arithmetic Properties of the values of the exponential function at algebraic points	45
3.1	Transcendence of e	45

3.2	Lindemann's theorem	52
3.3	The Gelfond-Schneider Theorem and Some Related Results	54
4	The computation of Transcendental Functions	57
4.1	Introduction	57

Chapter 1

Introduction

Although the ancient Greeks knew about the existence of irrational numbers, the theory of transcendental numbers is only about 150 years old. It was born in 1844 when Liouville established, for the first time, the existence of transcendental numbers. Since then, the recent advances of the theory, especially around Liouville's and hermite's theorems on linear forms in logarithms, have proved useful in many areas of number theory. This Report will focus on the algebraic and transcendence properties of some values, going into detail how some algebraic numbers are approximated by some other real numbers, for example rational, amongst others. To top it off, we shall look at 2 of the major forms of the transcendental numbers, the usual exponential, e ; and the pi, π . We will sketch the present state of knowledge on this topic and describe some of the tools that are involved in the proofs, addressing out open questions and potential avenues for further progress.

1.1 The pioneers in this field

But before we even go into the topic of transcendental numbers, it is only proper that we look at the forefathers of this field; without whom we would have never been able to encounter such wonders in the field of number theory.

Charles Hermite(1822-1901)

French mathematician who did brilliant work in many branches of mathematics, but was plagued by poor performance in exams as a student.

However, on his own, he mastered Lagrange's memoir on the solution of numerical equations and Gauss's *Disquisitiones Arithmeticae*.

Hermite did pioneering work on Abelian Functions, and he was the one who first proved that e was a transcendental number.

Joseph Liouville(1809-1882)

French mathematician who, with Sturm, developed many properties of boundary value problems. He also studied analysis and differential equations as well as demonstrating the existence of transcendental numbers, proving that the sum

$$\sum_{k=1}^{\infty} \frac{1}{n^{k!}} = \frac{1}{n} + \frac{1}{n^2} + \frac{1}{n^6} + \frac{1}{n^{24}} + \frac{1}{n^{120}} + \dots$$

is transcendental, where n is a real number greater than 1.

With $n = 10$, this is known as Liouville's number.

1.2 Some basic ideas

We shall now look at the different basic directions of research in the theory of transcendental numbers

Definition 1. A rational number [1] is a number that can be expressed in the form of a/b , where a and b are integers with $b > 0$

Theorem 1. A real number is a *rational* number if and only if it can be expressed as a repeating decimal, that is if and only if

$\alpha = m.d_1d_2 \cdots d_k \overline{d_{k+1}d_{k+2} \cdots d_{k+r}}$, where $m = [\alpha]$ if $\alpha \geq 0$ and $m = -[|\alpha|]$ if $\alpha < 0$, where k and r are non-negative integers with $r \geq 1$, and where d_j are digits.

Proof. If

$$\alpha = m.d_1d_2 \cdots d_k \overline{d_{k+1}d_{k+2} \cdots d_{k+r}}$$

then $(10^{k+r} - 10^k)\alpha \in \mathbb{Z}$ and it easily follows that α is rational.

If $\alpha = a/b$ with a and b both integers and $b > 0$, then $\alpha = m.d_1d_2 \cdots$ for some digit d_j . If $\{x\}$ denotes the fractional part of x , then

$$\{10^j|\alpha|\} = 0.d_{j+1}d_{j+2} \cdots \tag{1.1}$$

On the other hand,

$$\{10^j|\alpha|\} = \{10^j a/b\} = u/b \quad \text{for some } u \in \{0, 1, \dots, b-1\}.$$

Hence by the pigeon hole principle (which will be discussed in detail in the next chapter), there exists non-negative integer k and r with $r \geq 1$ and

$$\{10^k|\alpha|\} = \{10^{k+r}|\alpha|\}.$$

From (1.1), we deduce that

$$0.d_{k+1}d_{k+2}\cdots = 0.d_{k+r+1}d_{k+r+2}\cdots$$

so that

$$\alpha = m.d_1d_2\cdots d_k\overline{d_{k+1}d_{k+2}\cdots d_{k+r}}$$

And thus, Theorem 1 is proven.

Defintion 2. A number is irrational if it is not rational.

Theorem 2. A real number α which can be expressed as a non-repeating decimal is irrational.

Proof. From theorem 1 in the previous page, we can easily see that if $\alpha = m.d_1d_2\cdots$ and $\alpha = a/b$ is rational, then the digits d_j repeat. This will imply theorem 2 as if α does not have a repeating decimal, then we can invoke theorem 1 to show that it is not a rational number.

Defintion 3 A number α is said to be *algebraic* if it is a root of a polynomial

$$f(x) = a_nx^n + \cdots + a_1x + a_0, \quad f(x) \neq 0$$

with rational coefficients.

To prove that a given number α is algebraic, we need to find a non-zero polynomial $f(x) \in \mathbb{Q}[x]$ such that

$$f(\alpha) = 0$$

Definition 3 If the real number α is a root of

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{Z}[x]$$

then α is either an integer or an irrational number

Definition 4 A number which is not the root of *any* polynomial equation with integer coefficients, meaning that it's not an algebraic number of any degree, is said to be *transcendental*.

This definition guarantees that every transcendental number must also be irrational, since a rational number, by definition, is an algebraic number of degree 1.

1.3 Background Information

Transcendental numbers are important in the history of mathematics because of their investigation provided the first proofs that Circling of Square, one of the Geometric Problems of Antiquity, which had baffled mathematicians for more than 2000 years was, in fact, insoluble.

Specifically, in order for a number to be produced by a Geometric Construction using the ancient Greek rules, it must be either Rational or a Euclidean Number(a special case of algebraic numbers).

Because the number π is transcendental, the construction cannot be done according to the Greek rules.

Georg Cantor was the first to prove the existence of transcendental numbers. His proof will be given in chapter 2(section 3).

Liouville subsequently showed how to construct special cases(such as Liouville's constant), using Liouville's Rational Approximation Theorem(or Liouville's Theorem, in short).

In particular, he showed that any number, which has a rapidly converging sequence of rational approximations must be transcendental. Liouville also came up with the first rigorous proof of the existence of transcendental numbers in 1844.

It is in this light that we proceed into this project, to carry out investigations into the different properties of this special set of numbers.

1.4 Abstract

Now, we would give a brief outline on what we will be writing in the next few chapters:

1.4.1 Chapter 2

In this chapter, we will be discussing the many ways of approximation of different types of numbers;

these include

- Approximation of real numbers by algebraic numbers
- Simultaneous approximation
- Approximation of algebraic numbers by rational numbers (using Liouville's theorem)
- Approximation of algebraic numbers by algebraic numbers

The last section in chapter 2 will be dedicated to the refinements and generalizations of Liouville's theorem, and 3 of these refinements will be discussed, they are

- Thue's theorem
- Roth's theorem
- Thue-Siegel-Roth's theorem

1.4.2 Chapter 3

In this chapter, we will be looking in detail, the basic exponential function, e^x .

Fields of interest that we will be going into are:

- The different properties of the exponential-function using number theory

- Discussion of *Lindemann's theorem*
- Gelfond-Schneider's theorem(which is, in short, to proving that α^β is also transcendental)

1.4.3 Chapter 4

In this chapter, we will be looking at the different steps involved in obtaining a good computation in computer softwares for calculating transcendental functions.

We will be looking at a particular example, on how to calculate the exponential function, or e^x and give an outline on how the different steps are implemented.

Chapter 2

Approximation of real and algebraic numbers

2.1 Approximation of real numbers by algebraic numbers

We will be using the following notations in this report:

\mathbb{N} denotes the set of natural numbers,

\mathbb{Z} is the ring of rational integers,

\mathbb{Z}^+ is the set of nonnegative rational integers,

\mathbb{Q} is the field of rational numbers,

\mathbb{R} is the field of real numbers,

\mathbb{C} is the field of complex numbers,

and \mathbb{R}^m denotes the m -dimensional real euclidean space.

Let $\alpha \in \mathbb{R}$. For various $p \in \mathbb{Z}$ and $q \in \mathbb{N}$, we consider the modulus of the difference

$$\left| \alpha - \frac{p}{q} \right|. \quad (2.1)$$

Since the set of rational number is everywhere dense in the set of real numbers, it follows that for a suitable choice of p and q the magnitude of (2.1) can be made to be arbitrarily small. Thus, it makes sense to consider the relative smallness, ie how small we can make

it if q - the denominator of the fractional approximation- is not allowed to exceed some prescribed number.

Let $\varphi(q)$ be a function which is positive for all $q \in \mathbb{N}$. We consider the inequality

$$0 < \left| \alpha - \frac{p}{q} \right| < \varphi(q).$$

For a given α , it would be interesting to know:

for which functions $\varphi(q)$ does this inequality have an infinite set of solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$, and for which functions $\varphi(q)$ does the inequality have no solutions as soon as q exceeds some bound?

We say that a real number α has rational approximations p/q of order $\varphi(q)$ if there exists a constant $c > 0$ (which depends only on α and the function $\varphi(q)$) such that the inequality

$$0 < \left| \alpha - \frac{p}{q} \right| < c\varphi(q)$$

has infinitely many solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$.

The most common choice of function is a power function (as shown below):

$$\varphi(q) = \frac{1}{q^v}, \quad v > 0.$$

In that case, one is asking about the set of solutions of the inequality

$$0 < \left| \alpha - \frac{p}{q} \right| < c \frac{1}{q^v}$$

for different positive values of v and c .

Suppose that α is a rational number: $\alpha = a/b$, $(a, b) \in \mathbb{Z} \times \mathbb{N}$, $\text{GCD}(a, b) = 1$. for any $q \in \mathbb{N}$, there exist $p \in \mathbb{Z}$ such that

$$\frac{p}{q} < \frac{a}{b} \leq \frac{p+1}{q}.$$

Then

$$0 < \left| \alpha - \frac{p}{q} \right| \leq \frac{1}{q}. \tag{2.2}$$

If we let q take on different values, we find that there exists an infinite set of fractions p/q , p/q not equal to α which satisfies (2.2). Thus, α has rational approximations p/q of order $1/q$.

On the other hand, for any fraction p/q , p/q not equal to a/b , we have

$$\left| \alpha - \frac{p}{q} \right| = \left| \frac{a}{b} - \frac{p}{q} \right| = \frac{|aq - bp|}{bq} \geq \frac{1}{bq}. \quad (2.3)$$

It's clear to see that $|aq - bp| \geq 1$ as $p/q \neq a/b \Rightarrow (aq - bp) \neq 0$

From the inequality in (2.3), we find that, for any constant c with $0 < c \leq \frac{1}{b}$, the inequality

$$0 < \left| \alpha - \frac{p}{q} \right| < \frac{c}{q}$$

has no solution p/q .

This gives us the following theorem for irrationality.

Theorem 1 If $\varphi(q) > 0$ for all $q \in \mathbb{N}$, the inequality

$$0 < \left| \alpha - \frac{p}{q} \right| < \varphi(q)$$

has infinitely many solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$, and

$$\lim_{q \rightarrow \infty} q\varphi(q) = 0$$

then α is an irrational number.

Proof. comparing $\varphi(q)$ to that of the previous page, where $\varphi(q) = \frac{1}{q^v}$,

Hence, for $\left| \alpha - \frac{p}{q} \right| < \varphi(q)$, we have seen that for $\alpha \in$ rational numbers, there exists no solutions for p/q ,

Hence, if there exist solutions(infinitely many), α has to be a irrational number. This proves Theorem 1.

The order of approximation by rational numbers is different for different real numbers. we shall soon show that any irrational real number has rational approximations p/q of order $1/q^2$, and also there exist real numbers with "arbitrary high" order of approximation.

Dirichlet's Theorem If $\alpha \in \mathbb{R}$, $t \in \mathbb{R}$ and $t \geq 1$, then there exist $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ such that

$$|\alpha - \frac{p}{q}| < \frac{1}{qt}, \quad 0 < q \leq t. \quad (2.4)$$

Proof By Shidlovskii's Proof [2], let $\{a\}$ and $[a]$ denote the fractional part and integer part of a , respectively. We set $T = [t]+1$ and consider the $T+1$ numbers:

$$0 = \{\alpha \cdot 0\}, \{\alpha \cdot 1\}, \dots, \{\alpha(T-1)\}, 1 \quad (2.5)$$

Since $\{\alpha x\} = \alpha x - [\alpha x]$, $0 \leq \{\alpha x\} < 1$, $x = 0, 1, \dots, T-1$

All of the points in (2.5) belong to the interval of $0 \leq q \leq t$. We divide this interval into T equal parts of length $1/T$:

$$\frac{k}{T} \leq y < \frac{k+1}{T}, \quad k = 0, 1, \dots, T-2, \quad \frac{T-1}{T} \leq y \leq 1 \quad (2.6)$$

Each of the points in (2.5) lies in one of the interval in (2.6). Since the number of points in (2.5) is greater than the number of intervals in (2.6), it follows that there must be an interval which contains two of the points in (2.5). We now consider such a interval and there can be 2 possible cases.

Case 1. The interval is not the one at the extreme right. Suppose it contains the points $\{\alpha x_1\}$ and $\{\alpha x_2\}$, where $x_1 < x_2$. Then we have

$$|\{\alpha x_2\} - \{\alpha x_1\}| = |\alpha(x_2 - x_1) - [\alpha x_2] + [\alpha x_1]| < \frac{1}{T} < \frac{1}{t}.$$

We can now set $x_2 - x_1 = q$ and $[\alpha x_2] - [\alpha x_1] = p$. Obviously, $0 < q \leq T-1 \leq t$. We then have the inequalities

$$|\alpha q - p| < \frac{1}{t}, \quad 0 < q \leq t \quad (2.7)$$

from which (2.4) will follow.

Case 2. Our interval is the extreme right interval in (2.6). Suppose it contains the points $\{\alpha x_1\}$ as well as 1, where $x_1 \neq 0$. Then

$$|\{\alpha x_1\} - 1| = |\alpha x_1 - [\alpha x_1] - 1| \leq \frac{1}{T} < \frac{1}{t}$$

We can now set $x_1 = q$ and $[\alpha x_1] + 1 = p$ and then we can have $0 < q \leq T-1 \leq t$ and again we obtain (2.7), from which (2.4) follows.

Note If $t \in \mathbb{N}$ in the hypothesis of the theorem, then the proof is a little simpler: We can replace T by t , replace 1 by $\{ \alpha t \}$ in (2.5) and consider only the first case we discussed before.

The method used to prove Dirichlet's theorem is known as the *Dirichlet Pigeon Hole Principle*- *If m objects are placed in n boxes with $m > n$, then at least one box must contain two or more objects.*

Suppose that $\alpha \in \mathbb{Q}$, $\alpha = a/b$, $(a, b) \in \mathbb{Z} \times \mathbb{N}$ and $(a, b) = 1$. In this case we showed that the inequality (2.3) holds for any fraction not equal to α . Hence, if $t \geq b$, then (2.4) has only the trivial solution $p/q = a/b$. If $t < b$, then by Dirichlet's Theorem, the inequality (2.4) has a solution with denominator q satisfying $q \leq t < b$.

Hence, the denominators of all non-trivial solutions of (2.4) for all possible values of t are bounded, and so (2.4) has only a finite number of solutions (p, q) . Consequently, for $\alpha \in \mathbb{Q}$, the theorem gives us some information about approximations of a rational number by rational numbers with a smaller denominator.

Theorem 2 For any irrational number $\alpha \in \mathbb{R}$, the inequality

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^2}, \quad (2.8)$$

has infinitely many solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$, and hence the set of denominators of the solutions to (2.8) is unbounded.

Proof By Dirichlet's Theorem, for any solution (p, q) of (2.4) we have

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

But by what was proven before, (2.4) has infinitely many solutions (p, q) corresponding to different t . Consequently, the inequality (2.8) also has an infinite set of solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$. From this theorem, it follows that any irrational real number has approximations of order $\frac{1}{q^2}$.

From theorem 1, we know that any irrational number $\alpha \in \mathbb{R}$ can be represented in the

form of

$$\alpha = \frac{p}{q} + \frac{\theta_q}{q^2}, \quad (p, q) \in \mathbb{Z} \times \mathbb{N}, \quad |\theta_q| < 1 \quad (2.9)$$

where q is not an arbitrary natural number, but is a number which can be chosen to be arbitrary large, ie, there exists an infinite sequence of pairs $(p_n, q_n) \in \mathbb{Z} \times \mathbb{N}$, $q_{n+1} > q_n$ such that we can set $p = p_n$ and $q = q_n$ in (2.9).

The representation of an irrational real number in the form of (2.9) is often used in number theory and in other branches of pure and applied mathematics.

It is easy to construct examples of irrational numbers having rational number approximations of order greater than any desired power of q .

Example We consider the number:

$$\alpha = \sum_{k=1}^{\infty} \frac{1}{a^{k!}}, \quad a \in \mathbb{N}, \quad a \geq 2. \quad (2.10)$$

We denote

$$\sum_{k=1}^{\infty} \frac{1}{a^{k!}} = \frac{p_n}{q_n}, \quad q_n = a^{n!}, \quad p_n \in \mathbb{N}, \quad n = 1, 2, \dots$$

Then

$$\alpha - \frac{p_n}{q_n} = r_n = \sum_{k=1}^{\infty} \frac{1}{a^{k!}} > 0, \quad n = 1, 2, \dots$$

and

$$r_n = \frac{1}{a^{(n+1)!}} \left(1 + \frac{1}{a^{(n+2)! - (n+1)!}} + \dots \right) < \frac{1}{a^{(n+1)!}} \left(1 + \frac{1}{a} + \frac{1}{a^2} + \dots \right) = \frac{a}{a-1} \frac{1}{a^{(n+1)!}} = \frac{a}{a-1} \frac{1}{q_n^{n+1}} \leq \frac{1}{q_n^n}.$$

Hence

$$0 < \alpha - \frac{p_n}{q_n} < \frac{1}{q_n^n}, \quad n = 1, 2, \dots$$

The last inequality implies that for any non-negative real v , the inequality

$$0 < \left| \alpha - \frac{p_n}{q_n} \right| < \frac{1}{q_n^v}, \quad (2.11)$$

has a solution $(p, q) \in \mathbb{Z} \times \mathbb{N}$, $q \geq 2$. But any solution of (2.11) with a given value of v is also a solution of the same inequality for any smaller value of v . This implies that for any

$v \in \mathbb{R}, v > 0$, the inequality (2.11) has infinitely many solutions.

For $a = 10$, the series(2.10) reduces to the number with decimal expansion

$$\alpha = 0.110001000 \dots$$

which as the digit in the $(n!)$ -th place to the right of the decimal point and 0's everywhere else.

It is quite simple to construct examples of this sort using continued fractions. Moreover, it is not hard to prove the stronger result that there exist irrational real numbers having rational approximations which are better than any given function of q .

Theorem 3 Given any positive function $\varphi(q)$ on the natural numbers, there exists an irrational $\alpha \in \mathbb{R}$ such that the inequality

$$\left| \alpha - \frac{p}{q} \right| < \varphi(q)$$

has infinitely many solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$.

Proof. We define a sequence $\{ \gamma_n \}$, $\gamma_n \in \mathbb{R}, \gamma_n \geq 0$, in such a way that the sequence of natural numbers

$$m_1 = 1, \quad m_n = \left[\log_2 \frac{1}{\varphi^{2^{m_{n-1}}}} + \gamma_{n-1} \right] + 1, n = 2, 3, \dots$$

satisfies the condition $m_{n+1} > 2m_n$. We then set(define)

$$\alpha = \sum_{k=1}^{\infty} \frac{(-1)^k}{2^{m_k}} \tag{2.12}$$

and

$$q_n = 2^{m_n}, \quad p_n = q_n \sum_{k=1}^{\infty} \frac{(-1)^k}{2^{m_k}}, n = 1, 2, \dots$$

Since (2.12) is an alternating series with decreasing terms, we have the inequalities

$$\begin{aligned} 0 < \frac{1}{2^{m_{n+1}}} - \frac{1}{2^{m_{n+2}}} < \left| \alpha - \frac{p}{q} \right| < \frac{1}{2^{m_{n+1}}} \\ &< \frac{1}{2^{\log_2 \frac{1}{\varphi^{2^{m_n}}}}}. \end{aligned}$$

By a simple substitution, ie letting $e = 2^{\log_2 \frac{1}{\varphi^{2^{m_n}}}}$, we will get

$$\frac{1}{2^{\log_2 \frac{1}{\varphi^{2^{m_n}}}}} = \varphi(2^{m_n}) = \varphi(q_n). \quad (2.13)$$

Since $m_{n+1} > 2m_n$, it follows from (3.13) that

$$0 < \left| \alpha - \frac{p}{q} \right| < \frac{1}{2^{m_{n+1}}} < \frac{1}{2^{2m_n}} = \frac{1}{q_n^2}.$$

Now α is irrational by making use of Theorem 1, and then Theorem 3 follows from (2.13)

The set of real numbers having very good rational approximation is relatively small. Almost all real numbers are not approximated very well by rational numbers. This follows from a proof by Khinchin in 1924, which will be stated here, but not proven.

Khinchin's Theorem Let $f(x)$ be a positive continuous function for $x \geq c$ where $c > 0$ such that $xf(x)$ is non-increasing. Then the inequality

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

has an infinite set of solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$ for almost all α provided that the integral

$$\int_c^\infty f(x) dx$$

diverges; while if this integral converges, for almost all α , the above inequality has only finitely many solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$.

This theorem implies that the inequality

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^2 \ln q}$$

has an infinite set of solutions for almost all α .

Remarks Dirichlet's Theorem can be proven in other ways, for example using continued fractions or farey series. But the simplest method of proof is to use the Dirichlet Pigeon-Hole Principle. The Pigeon-Hole Principle can be applied to many situations in the theory

of Diophantine approximation when continued fractions or Farey series cannot be used, such as certain theorems on simultaneous approximation of a set of numbers, and also in many proofs in the theory of transcendental numbers. Some examples of such problems will be examined in the next section.

2.2 Simultaneous Approximation

We now consider a linear form of several real variables with integer coefficients bounded by some number and not all zero. We shall discuss how small it is possible to make the modulus of this form by suitably choosing the coefficients.

We shall now state and prove a theorem similar to that of Theorem 1 from the previous chapter,

Theorem 4 (Dirichlet) If $\alpha_1, \dots, \alpha_m$ are real numbers, $m \geq 1$, and if $t \in \mathbb{R}$, $t \geq 1$, then there exist $a_1, \dots, a_m, b \in \mathbb{Z}$ such that

$$|a_1\alpha_1 + \dots + a_m\alpha_m - b| < \frac{1}{t^m}, 0 < \max|a_k| \leq t. \quad (2.14)$$

Proof We set $T = [t] + 1$ and consider the following $T^m + 1$ numbers: T^m fractional parts

$$\gamma = \{a_1\alpha_1 + \dots + a_m\alpha_m\}, 0 \leq \gamma < 1, \quad (2.15)$$

where each of the a_1, \dots independently runs through the T values $0, 1, \dots, T-1$, and also the number 1. Now we use the methods that is used in the previous section (ie, the Pigeon-Hole Principle).

We divide the interval $0 \leq y \leq 1$ into T^m parts of length $\frac{1}{T^m}$:

$$\frac{k}{T^m} \leq y < \frac{k+1}{T^m}, \quad k = 0, 1, \dots, T^m - 2, \quad \frac{T^m - 1}{T^m} \leq y \leq 1. \quad (2.16)$$

The points in (2.15) along with 1 all belong to the interval $0 \leq y \leq 1$. Hence, these $T^m + 1$ points each fall in one of the T^m intervals (2.16). Since the number of points is greater than the number of intervals, there must be an interval containing at least 2 points. Two cases are possible.

1. The interval containing two of our points is not the one on the extreme right. Suppose that this interval contains the points $\gamma' = \{a_1'\alpha_1 + \dots + a_m'\alpha_m\}$ and $\gamma'' = \{a_1''\alpha_1 + \dots + a_m''\alpha_m\}$. Then

$$\begin{aligned} & |\{a_1''\alpha_1 + \dots + a_m''\alpha_m\} - \{a_1'\alpha_1 + \dots + a_m'\alpha_m\}| \\ &= |a_1\alpha_1 + \dots + a_m\alpha_m - b| < \frac{1}{T^m} = \frac{1}{([t] + 1)^m} < \frac{1}{t^m} \end{aligned} \quad (2.17)$$

where

$$\begin{aligned} a_k &= a_k'' - a_k', \quad k = 1, \dots, m; \\ b &= [a_1'' \alpha_1 + \dots + a_m'' \alpha_m] - [a_1' \alpha_1 + \dots + a_m' \alpha_m], \end{aligned} \quad (2.18)$$

in which $0 < \max |a_k| \leq T-1 = [t] \leq t$.

2. Our interval is the one at the extreme right. Suppose that it contain the points

$$\gamma = \{a_1 \alpha_1 + \dots + a_m \alpha_m\}, \quad 0 < \max |a_k| \leq T - 1 \leq t \quad (2.19)$$

and 1. Then

$$|\{a_1 \alpha_1 + \dots + a_m \alpha_m\} - 1| = |a_1 \alpha_1 + \dots + a_m \alpha_m - b| \leq \frac{1}{T^m} < \frac{1}{t^m} \quad (2.20)$$

where $b = [a_1 \alpha_1 + \dots + a_m \alpha_m] + 1$.

The inequality (2.14) follows from (2.17) and (2.18) in the first case and from (2.19) and (2.20) in the second case.

This proof is completely the analogous to the proof of Dirichlet's Theorem in section 1; in fact the latter theorem follows from theorem 4 by letting $m = 1$.

We now prove a theorem on simultaneous approximation of a set of numbers

Kronecker's Theorem If $\alpha_1, \dots, \alpha_m$ are arbitrary real numbers, $m \geq 1$, and if $t \in \mathbb{N}$, then there exist $p_1, \dots, p_m \in \mathbb{Z}$ and $q \in \mathbb{N}$ such that

$$|\alpha_k - \frac{p_k}{q}| < \frac{1}{qt}, \quad k = 1, \dots, m, \quad 0 < q \leq t^m. \quad (2.21)$$

Proof In \mathbb{R}^m we consider the points

$$M_x(\{\alpha_1(x)\}, \dots, \{\alpha_m(x)\}), \quad x = 0, 1, \dots, t^m, \quad (2.22)$$

Where $\{\alpha_k(x)\} = \alpha_k(x) - [\alpha_k(x)]$, $0 \leq \{\alpha_k(x)\} < 1$, $k = 1, \dots, m$ and we consider the unit cube, ie the set of points

$$M(y_1, \dots, y_m), \quad 0 \leq y_k < 1, \quad k = 1, 2, \dots, m. \quad (2.23)$$

We divide the segment $0 \leq y_i < 1$ of each coordinate axis into t parts of length $1/t$ and we use the corresponding hyperplanes to divide the unit cube into t^m small cubes, each of which consists of all points $Q(y_1, \dots, y_m)$ whose coordinates satisfy the inequalities

$$\frac{k_1}{t} \leq y_1 < \frac{k_1 + 1}{t}, \dots, \frac{k_m}{t} \leq y_m < \frac{k_m + 1}{t}, \quad (2.24)$$

Where the k_i , $i = 1, \dots, m$ are an m -tuple of integers from $0, 1, \dots, t-1$. All of the t^m+1 points in (2.22) are contained in the unit cube (2.23). Each of these points fall in exactly one of the small cubes (2.24). Hence, there is some small cube which contain at least two of the small points in (2.22), let's say

$$M_{x_1}(\{\alpha_1 x_1\}, \dots, \{\alpha_m x_1\}), M_{x_2}(\{\alpha_1 x_2\}, \dots, \{\alpha_m x_2\}), \quad x_2 > x_1$$

Since these two points lie in the same cube, it follows that the absolute value of the difference between any two corresponding coordinates is less than $1/t$. That is,

$$|\{\alpha_k x_2\} - \{\alpha_k x_1\}| < \frac{1}{t}, \quad k = 1, \dots, m;$$

in other words, if we set $q = x_2 - x_1$ and $p_k = [\alpha_k x_2] - [\alpha_k x_1]$ for $k = 1, \dots, m$, then we have

$$|\alpha_k q - p_k| < \frac{1}{t}, \quad k = 1, 2, \dots, m, \quad 0 < q \leq t^m. \quad (2.25)$$

These inequalities imply (2.21). Similar for the previous case, when $m=1$, Kronecker's Theorem becomes Dirichlet's Theorem (with $t \in \mathbb{N}$). Just as in the case of Dirichlet's Theorem, from Kronecker's Theorem, it is easy to obtain the following

Corollary Under the assumptions in Kronecker's Theorem, if at least one of the $\alpha_1, \dots, \alpha_m$ is irrational, then the inequalities

$$|\alpha_k - \frac{p_k}{q}| < q^{-1-\frac{1}{m}}, \quad k = 1, 2, \dots, m, \quad 0 < q \leq t^m.$$

have an infinite set of solutions $p_1, \dots, p_m \in \mathbb{Z}$, $q \in \mathbb{N}$. We next prove a theorem giving a bound for linear forms which we shall need later.

Theorem 5 If $\alpha_1, \dots, \alpha_m$ are arbitrary complex numbers, $m \geq 2$, and if $H \in \mathbb{N}$ then there exist numbers

$$a_k \in \mathbb{Z}, \quad |a_k| \leq H, \quad k = 1, 2, \dots, m \quad \max |a_k| > 0 \quad (2.26)$$

such that the linear form

$$L = a_1 \alpha_1 + \dots + a_m \alpha_m \quad (2.27)$$

satisfies the inequality

$$|L| \leq cH^{1-rm}, \quad (2.28)$$

where $r = 1$ if all of the α_k are real and $r = 1/2$ if one or more of them is complex, and

$$c = \begin{cases} \sum_{k=1}^m |\alpha_k| & \text{if } r = 1 \\ \sqrt{2} \sum_{k=1}^m |\alpha_k| & \text{if } r = 1/2 \end{cases}$$

Proof. The theorem is obvious if $H=1$; so we suppose that $H \geq 2$.

We consider all possible linear forms of the form (2.27), where α_k independently run through all integer values satisfying the inequalities

$$|a_k| \leq \left[\frac{H}{2}\right], \quad k = 1, \dots, m.$$

The number of such forms L is equal to

$$\left(2\left[\frac{H}{2}\right] + 1\right)^m. \quad (2.29)$$

Their values satisfy the inequalities

$$|L| \leq \gamma \left[\frac{H}{2}\right], \quad \gamma = \sum_{k=1}^m |\alpha_k|. \quad (2.30)$$

Explanation for (2.30) Since $|a_k| \leq \left[\frac{H}{2}\right]$, from

$L = a_1\alpha_1 + \dots + a_m\alpha_m$, we obtain

$$L \leq \left[\frac{H}{2}\right] (\alpha_1 + \dots + \alpha_m) \leq \left[\frac{H}{2}\right] \leq \left[\frac{H}{2}\right] \gamma, \text{ where } \gamma = \sum_{k=1}^m |\alpha_k|.$$

(This complete the proof for (2.30))

We consider two cases (we exclude the obvious case $\gamma = 0$)

Case 1. All of the $\alpha_1, \dots, \alpha_m$ are real numbers. Then the values of all of our forms L are contained in the interval with the endpoints $[-\gamma\left[\frac{H}{2}\right], +\gamma\left[\frac{H}{2}\right]]$, which has length $2\gamma\left[\frac{H}{2}\right]$. We divide this interval into

$$\left(2\left[\frac{H}{2}\right] + 1\right)^m - 1 \quad (2.31)$$

equal subintervals. The number of forms L (see (2.29)) is greater than the number (2.31) of subintervals. Hence, there exists a subinterval containing the values of two different forms L . Let these two forms be

$$L' = a_1' \alpha_1 + \dots + a_m' \alpha_m, \quad L'' = a_1'' \alpha_1 + \dots + a_m'' \alpha_m. \quad (2.32)$$

Then

$$|L' - L''| \leq 2 \frac{\gamma[\frac{H}{2}]}{(2[\frac{H}{2}] + 1)^m - 1}.$$

If H is an even number, then $2[\frac{H}{2}] = H$, and

$$|L' - L''| \leq \frac{\gamma H}{(H + 1)^m - 1} < \gamma H^{1-m}.$$

If H is an odd number, then $2[\frac{H}{2}] = H - 1$, and

$$|L' - L''| \leq \frac{\gamma H - 1}{(H)^m - 1} < \gamma H^{1-m}.$$

Thus,

$$|L^I - L^{II}| < \gamma H^{1-m}.$$

We set $a_k = a_k' - a_k''$, $k = 1, \dots, m$. Then the form

$$L = L' - L'' = a_1 \alpha_1 + \dots + a_m \alpha_m$$

satisfy the conditions

$$|a_k| \leq 2[\frac{H}{2}] \leq H, \quad \sum_{k=1}^m |\alpha_k| > 0$$

and

$$|L| < c H^{1-m}, \quad c = \gamma.$$

Case 2. At least one of the numbers a_1, \dots, a_m is complex. In this case, it follows from (2.30) that the values of the form L lie in a square centered at the origin with the sides parallel to the coordinate axes of length equal to $2\gamma[\frac{H}{2}]$. We then divide the sides of this square into M equal segments, where M is the integer satisfying the inequalities

$$(2[\frac{H}{2}] + 1)^{\frac{m}{2}} - 1 \leq M < (2[\frac{H}{2}] + 1)^{\frac{m}{2}}. \quad (2.33)$$

Using lines through the division points parallel to the coordinate axes, we divide our square into M^2 small square. From (2.33), we have

$$M^2 < (2[\frac{H}{2}] + 1)^m.$$

This inequality means that the number of small squares is less than the number (2.29) of linear forms L . Hence there exists a small square which contains the values of two different forms L . Suppose that these two forms are the ones in (2.32). We then find that $|L' - L''|$ is no greater than the diagonal of a small square. This observation, together with (2.33), enables us to say that the following inequalities hold for any $H \geq 2$ and $m \geq 2$:

$$|L' - L''| \leq \frac{\sqrt{2}\gamma[\frac{H}{2}]}{M} \leq \frac{\sqrt{2}\gamma[\frac{H}{2}]}{(2[\frac{H}{2}] + 1)^{\frac{m}{2}} - 1}. \quad (2.34)$$

If H is an even number, then $2[H/2] = H$ and since $m \geq 2$, then we can conclude that from (2.34) that

$$|L' - L''| \leq \frac{\sqrt{2}\gamma H}{(H + 1)^{\frac{m}{2}} - 1} \leq \sqrt{2}\gamma H^{1-\frac{m}{2}}.$$

If H is an odd number, then $2[H/2] = H-1$, and we can similarly conclude that from (2.34)

$$|L' - L''| \leq \frac{\sqrt{2}\gamma H - 1}{(H)^{\frac{m}{2}} - 1} = \sqrt{2}\gamma(H)^{1-\frac{m}{2}} \frac{1 - H^{-1}}{1 - H^{-m/2}} \leq \sqrt{2}\gamma(H)^{1-\frac{m}{2}}.$$

We can then argue along the same line as the first case in order to obtain the inequality (2.28) with $c = \sqrt{2}\gamma$.

2.3 Approximation of algebraic numbers by rational numbers

The arithmetic operations applied to algebraic numbers give us algebraic numbers again. Hence, the set of all algebraic numbers forms a field, which we shall denote \mathbb{A}

Let α be an algebraic number. There exists a unique irreducible polynomial $f(x) \in \mathbb{Q}[x]$ with leading coefficient 1 which has α as a root. The polynomial is called the minimal polynomial of α , and its degree is called the degree of α , denoted by $\deg(\alpha)$. If α is an algebraic number of degree n , then the roots $\alpha = \alpha_1, \alpha_2, \dots, \alpha_n$ of the minimum polynomial of α are called the *conjugates* of α . We set

$$|\overline{\alpha}| = \max|\alpha_k|$$

and call this number the size of the algebraic number α . We can now prove the Liouville's Theorem on the approximation of an algebraic number by rational numbers.

Liouville's Theorem If α is a real algebraic number of degree n , $n \geq 1$, then there exists a constant $c = c(\alpha) > 0$ such that the following inequality holds for any $p \in \mathbb{Z}$ and $q \in \mathbb{N}$, p/q not equal to α :

$$\left| \alpha - \frac{p}{q} \right| > \frac{c}{q^n}. \quad (2.35)$$

Proof Suppose that α is a root of the irreducible polynomial

$$f(x) = a_n x^n + \dots + a_1 x + a_0, \quad f(x) \in \mathbb{Z}[x], a_n > 0$$

and $\alpha = \alpha_1, \dots, \alpha_n$ are conjugates of α . Then by definition of α being a root

$$|f(x)| = a_n |x - \alpha| \prod_{k=2}^n |x - \alpha_k|. \quad (2.36)$$

Let p and q be arbitrary integers with $q > 0$ and $p/q \neq \alpha$. We consider two cases.

Case 1. If $n = 1$, we have

$$f(x) = a_1 x + a_0;$$

$$\begin{aligned}
0 &= f(\alpha) = a_1\alpha + a_0 = 0; \\
&\Rightarrow \alpha = -\frac{a_0}{a_1}.
\end{aligned}$$

Then

$$|\alpha - \frac{p}{q}| \geq \frac{1}{a_1q} = \frac{c}{q}$$

Case 2. Let $n > 1$. If p and q are such that $|\alpha - \frac{p}{q}| \geq 1$, then

$$|\alpha - \frac{p}{q}| \geq \frac{1}{q^n} \tag{2.37}$$

On the other hand, if $|\alpha - \frac{p}{q}| < 1$, then we have

$$|\frac{p}{q}| < |\alpha| + 1 \leq \overline{|\alpha|} + 1. \tag{2.38}$$

In addition, $f(p/q) \neq 0$, since the irreducible polynomial $f(x)$ cannot have a rational root for $n > 1$. We now set $x = p/q$ in (2.36). If we note that

$$|f(\frac{p}{q})| = \frac{|a_n p^n + \dots + a_0 q^n|}{q^n} \geq \frac{1}{q^n}$$

and by (2.38),

$$\prod_{k=2}^n |\frac{p}{q} - \alpha_k| \leq \prod_{k=2}^n (|\frac{p}{q}| + |\alpha_k|) < (2\overline{|\alpha|} + 1)^{n-1}$$

But we know that

$$|\frac{p}{q}| \leq \overline{|\alpha|} + 1 \text{ (From (2.38))}$$

Thus

$$|\frac{p}{q}| + |\alpha_k| \leq \overline{|\alpha|} + 1 + |\alpha_k| < \overline{|\alpha|} + 1 + \overline{|\alpha|} \text{ (as } \overline{|\alpha|} = \max|\alpha|) < (2\overline{|\alpha|} + 1)^{n-1}$$

then we obtain

$$\frac{1}{q^n} < a_n (2\overline{|\alpha|} + 1)^{n-1} |\alpha - \frac{p}{q}| \tag{2.39}$$

Now the inequality (2.35) follows from (2.37) and (2.39), if we set

$$c = c(\alpha) = \frac{1}{a_n (2\overline{|\alpha|} + 1)^{n-1}}$$

Hence the proof of Liouville's Theorem is completed.

Now, we give a corollary to Liouville's Theorem.

Corollary Under the conditions of Liouville's Theorem, there exists a constant $c=c(\alpha) >0$ such that the inequality

$$0 < \left| \alpha - \frac{p}{q} \right| < \frac{c}{q^n}$$

has no solution $(p, q) \in \mathbb{Z} \times \mathbb{N}$. What this means is that an algebraic number α of degree n does not have approximations by rational numbers of order better than $1/q^n$. Liouville's Theorem provides a necessary condition for a number α to be algebraic; in other words, it gives us a sufficient condition for transcendence. We shall state this as a theorem.

Theorem 6 Suppose that α is a real number such that for any $v \in \mathbb{R}$, $v > 0$, the inequality

$$0 < \left| \alpha - \frac{p}{q} \right| < \frac{1}{q^v} \tag{2.40}$$

has an infinite set of solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$. Then α is transcendental.

Proof: Suppose that α were an algebraic number of degree n . Using Liouville's Theorem, there would exist a constant $c(\alpha) > 0$ such that for any $(p, q) \in \mathbb{Z} \times \mathbb{N}$, $p/q \neq \alpha$, we have

$$\left| \alpha - \frac{p}{q} \right| > \frac{c\alpha}{q^n}. \tag{2.41}$$

If we set $v = n+1$ in (2.40) and choose a solution for $1/q < c(\alpha)$. Then

$$0 < \left| \alpha - \frac{p}{q} \right| < \frac{1}{q^{n+1}} < \frac{c(\alpha)}{q^n},$$

But this will contradict our original condition given in (2.41) and hence, we have proven the theorem.

Theorem 7 *Transcendental Numbers exist*

2 proofs will be given here:

Proof 1: gives a class of transcendental numbers

Proof 2: which gives a concrete example to prove the existence of transcendental numbers.

Proof 1:

The essence of this proof is that the real algebraic numbers are countable whereas the

set of all real numbers are uncountable, so there must exist real transcendental numbers.

Define

$$P(n) = \left\{ f(x) = \sum_{j=0}^n a_j x^j \in \mathbb{Z}[x] : 1 \leq \sum_{j=0}^n |a_j| \leq n \right\}.$$

Observe that $P(n)$ is finite. Also, every non-zero polynomial in $\mathbb{Z}[x]$ belongs to some $P(n)$. By considering the real root of polynomials in $P(1), P(2), \dots$ (at the k -th stage, consider the real roots of polynomials $P(k)$ which have not occurred as a root of a polynomial in $P(j)$ for $j > k$), we can order the algebraic numbers; hence they are countable. Next, Observe that the set of real numbers is clearly uncountable. This will complete the proof.

For Liouville's proof, we first establish a few results [1] needed for this proof,

Definition 1 A real number α is a Liouville number [5] if for every positive integer n , there are integers a and b with $b > 1$ such that

$$0 < \left| \alpha - \frac{a}{b} \right| < \frac{1}{b^n}.$$

We shall show that Liouville numbers exist. *Proof 2* of Theorem 7 will then follow from our next result.

Theorem 1. All Liouville numbers are *transcendental*.

Lemma 1. Let α be an irrational number which is a root $f(x) = \sum_{j=0}^n a_j x^j \in \mathbb{Z}[x]$ with $f(x) \neq 0$. Then there is a constant $A = A(\alpha) > 0$ such that if a and b are integers with $b > 0$, then

$$\left| \alpha - \frac{a}{b} \right| > \frac{A}{b^n} \quad (*)$$

Proof. Let M be the maximum value of $|f'(x)|$ on $[\alpha - 1, \alpha + 1]$. Let $\alpha_1, \alpha_2, \dots, \alpha_m$ be the distinct roots of $f(x)$ which are different from α . Fix

$$A < \min\left\{1, \frac{1}{M}, |\alpha - \alpha_1|, \dots, |\alpha - \alpha_m|\right\}.$$

Assume that (*) does not hold for some a and b integers with $b > 0$. Then

$$\left| \alpha - \frac{a}{b} \right| \leq \frac{A}{b^n} \leq A < \min\{1, |\alpha - \alpha_1|, \dots, |\alpha - \alpha_m|\}$$

Hence,

$$\frac{a}{b} \in [\alpha - 1, \alpha + 1] \quad \text{and} \quad \frac{a}{b} \notin \{\alpha_1, \dots, \alpha_m\}$$

By mean value theorem, there is an x_0 between a/b and α such that

$$f(\alpha) - f(a/b) = (\alpha - a/b)f'(x_0)$$

So that

$$\left| \alpha - \frac{a}{b} \right| = \left| \frac{f(a/b) - f(\alpha)}{f'(x_0)} \right| = \left| \frac{f(a/b)}{f'(x_0)} \right|.$$

Since $f(a/b)$ is not zero, we can deduce that

$$|f(a/b)| = \left| \sum_{j=0}^n a_j a^j b^{n-j} \right| / b^n \geq 1/b^n.$$

Thus, since $|f'(x_0)| \leq M$, we obtain

$$\left| \alpha - \frac{a}{b} \right| \geq \frac{1}{Mb^n} > \frac{A}{b^n} \geq \left| \alpha - \frac{a}{b} \right|.$$

Thus giving a contradiction. Thus, Lemma 1 is proven to be true.

Proof of Theorem 1: All Liouville number are transcendental. Let α be a Liouville number.

First, we show that α must be irrational. Assume $\alpha = c/d$ for some integer c and d with $d > 0$. Let n be a positive integer with $2^{n-1} > d$. Then for any integers a and b with $b > 1$ and $a/b \neq c/d$, we have

$$\left| \alpha - \frac{a}{b} \right| = \left| \frac{c}{d} - \frac{a}{b} \right| \geq \frac{1}{bd} > \frac{1}{2^{n-1}b} \geq \frac{1}{b^n}.$$

It follows from the definition of a Liouville number (ie using Liouville's Theorem) that α is NOT a liouville number, giving a contradiction. Thus α is irrational.

Now we assume α is an irrational algebraic number. By the lemma, there exist a real number $A > 0$ and a positive integer n such that (*) holds for all integers a and b with $b > 0$. Let r be a positive integer for which $2^r \geq 1/A$. Since α is a Liouville number, there are integers a and b with $b > 1$ such that

$$\left| \alpha - \frac{a}{b} \right| < \frac{1}{b^{n+r}} \leq \frac{1}{2^r b^n} \leq \frac{A}{b^n}.$$

This will contradict (*) and hence, establishes that α is transcendental.

We shall now give the 2nd proof for Theorem 7

We shall do so by giving a example(concrete example)

Proof 2 We show that $\alpha = \sum_{j=0}^{\infty} 1/2^{j!}$ is a Liouville number.

First, observe that the binary expansion of α has arbitrarily long strings of 0's and so it cannot be rational. Fix a positive integer n and consider $a/b = \sum_{j=0}^n 1/2^{j!}$ with a and $b = 2^{n!} > 1$ integers. Then

$$0 < \left| \alpha - \frac{a}{b} \right| = \sum_{j=n+1}^{\infty} \frac{1}{2^{j!}} < \sum_{j=(n+1)!}^{\infty} \frac{1}{2^j} \leq \frac{1}{2^{(n+1)!-1}} \leq \frac{1}{2^{n(n)!}} = \frac{1}{b^n}.$$

This proves that α is Liouville.

2.4 Approximation of algebraic numbers by algebraic numbers

By $H = H_\psi$ of a polynomial $\psi = \psi(x) \in \mathbb{C}[x]$, we mean the maximum modulus of its coefficients.

In general

$$H(b_n x^n + \dots + b_1 x + b_0) = \max(|b_n|, \dots, |b_1|, |b_0|)$$

Now suppose that $f(x)$ is the minimal polynomial of an algebraic number α . If we multiply it by the least common denominator of its coefficients, we obtain a primitive polynomial $\gamma(x) \in \mathbb{Z}[x]$ which also has α as a root (Recall that a polynomial is called *primitive* if there is no integer greater than 1 which divides all of its coefficients.) Then the height $H = H_\psi$ of the algebraic number α is defined to be the height of the irreducible primitive polynomial $\gamma(x) \in \mathbb{Z}[x]$ which has α as a root.

An algebraic number α is called an algebraic integer if all of the coefficients of its minimal polynomial $f(x)$ are rational integers.

The sum, difference or product of two algebraic integers is an algebraic integer. Consequently, the set of all algebraic integers form a ring, which we shall denote as \mathbb{Z}_A . If $\alpha \in A$, then there exists an $r \in \mathbb{N}$ such that $r\alpha \in \mathbb{Z}_A$.

To prove the above statement, we first note that there exist

$$\psi(x) = a_n x^n + \dots + a_0 \in \mathbb{Z}[x]$$

such that

$$\begin{aligned} \psi(\alpha) &= 0, a_n > 0 \\ \Rightarrow \psi(\alpha) &= a_n \alpha^n + \dots + a_0 = 0 \\ \Rightarrow a_n^n \alpha^n &+ \dots + a_n^{n-1} a_0 = 0 \end{aligned}$$

Rewriting, we get

$$\Rightarrow (a_n \alpha)^n + \dots + b_0 = 0, b_0, \dots \in \mathbb{Z} \Rightarrow a_n(\alpha) \in \mathbb{Z}_A.$$

So we simply take $r = a_n$)

If α is a root of $\psi(x) \in \mathbb{Z}[x]$, $\psi(x)$ not equivalent to 0, then we can choose r to be the modulus of the leading coefficient of $\psi(x)$.

By an algebraic number field, we mean the extension field of $\mathbb{Q}(\theta)$, where $\mathbb{Q}(x)$ runs through all rational functions in $\mathbb{Q}(x)$ whose denominator does not vanish at θ . Such a field will be denoted $\mathbb{Q}(\theta)$. θ is called a generating element of $\mathbb{Q}(\theta)$.

Different elements of this field can be chosen to be the generating element.

With $Q(\theta) = \{ \frac{p(\theta)}{q(\theta)} \mid p(x) \text{ and } q(x) \in \mathbb{Q}[x], q(\theta) \text{ non-zero.} \}$,

we can easily check that $Q(\theta)$ is a field

All of these generating elements have the same degree h .

Generating elements are called the *primitive* elements of the algebraic number field.

By the degree of the algebraic number field $K = \mathbb{Q}(\theta)$ we mean the degree of any generating element; this degree is denoted $[K : \mathbb{Q}] = \mathbb{Q}(\theta) : \mathbb{Q}$.

The degree h of an algebraic number field $\mathbb{Q}(\theta)$ is equal to the maximum number of linearly independent elements over \mathbb{Q} . Suppose that $K = \mathbb{Q}(\theta)$ is an algebraic number field, $[K : \mathbb{Q}] = h$, and $\theta = \theta_1, \dots, \theta_h$ are the conjugates of θ . Every element $\alpha \in K$ can be represented uniquely in the form of

$$\alpha = r(\theta) = c_0 + c_1\theta + \dots + c_{h-1}\theta^{h-1}, \quad c_i \in \mathbb{Q}, \quad i = 0, 1, \dots, h-1.$$

Claim: $(1, \theta, \dots, \theta^{h-1})$ is a basis of K over \mathbb{Q}

Proof:

1) Linear independence over \mathbb{Q} .

Let $f(x)$ be the polynomial $c_0 + c_1x + \dots + c_{h-1}x^{h-1} \in \mathbb{Q}[x]$

so $f(\theta) = 0$,

\Rightarrow Degree of $f(x) \leq h <$ Degree of min polynomial. Hence $f(x) = 0$;

Hence, in conclusion, we find that $f(x) = 0 \Rightarrow$ all its coefficients $c_0, c_1, \dots, c_{h-1} = 0$

Therefore, the set $\{1, \theta, \dots, \theta^{h-1}\}$ are linearly independent over \mathbb{Q}

2) Spanning of K .

We take $\frac{a(\theta)}{b(\theta)} \in K$

(where $b(\theta)$ not equal 0)

Suppose that $c(x) \mid b(x)$ and $c(x) \mid p(x)$, $c(x)$ not constant polynomial,
Therefore,

$$\begin{aligned} \lambda p(x) &\mid b(x) \\ \Rightarrow p(x) &\mid b(x) \end{aligned}$$

$$b(x) = p(x)r(x) \Rightarrow b(\theta) = p(\theta)r(\theta) = 0 \Rightarrow b(\theta) = 0, \text{ with } r(x) \in \mathbb{Q}[x].$$

We will get a contradiction here as $b(\theta)$ is given to be non-zero,
so $c(x)$ must be constant and the GCD of $b(x)$ and $p(x)$ is 1, Then by Euclid's algorithm,
we get

$$\lambda(x)b(x) + M(x)p(x) = 1.$$

If we set $x = \theta$, then

$$\begin{aligned} \lambda(\theta)b(\theta) + M(\theta)p(\theta) &= 1 \\ \Rightarrow \lambda(\theta) &= \frac{1}{b(\theta)} (\text{Recall } p(\theta) = 0) \\ \Rightarrow \frac{a(\theta)}{b(\theta)} &= a(\theta)\lambda(\theta) = \text{polynomial in } \theta \end{aligned}$$

Therefore, we let

$$c(\theta) = \frac{a(\theta)}{b(\theta)}$$

Since $p(\theta) = 0$, $c(\theta) = a_0 + \dots + a_{h-1}\theta^{h-1}$ for some $a_0, \dots, a_{h-1} \in \mathbb{Q}$.

Hence, K will be spanned by the set $\{1, \theta, \dots, \theta^{h-1}\}$

And this will complete the proof for showing that the set $\{1, \theta, \dots, \theta^{h-1}\}$ is a basis.

Next, we define the numbers

$$\alpha_i = r(\theta_i), i = 1, \dots, h, \tag{2.42}$$

where $r(\theta) = c_0 + c_1\theta + \dots + c_{h-1}\theta^{h-1}$ and the α_i are called the conjugates of α in the field K . They are the conjugates of α (in the sense defined before, ie, the roots of the minimal polynomial), but each conjugate is repeated $h/\deg(\alpha)$ times.

When we are considering an algebraic number field $K = \mathbb{Q}(\theta)$, we shall always suppose that the number $\theta = \theta_1, \dots, \theta_h$ are fixed. According to (2.42), we thereby also fix the numbering of the conjugates of any $\alpha \in K$.

If $\alpha_1, \dots, \alpha_h$ are the conjugates of α in K , and if $Q(x) \in \mathbb{Q}(x)$ is such that $Q(\alpha)$ is defined, then the elements $Q(\alpha_1), \dots, Q(\alpha_h)$ are the conjugates of $Q(\alpha)$ in K . Given any finite set of algebraic numbers, $\alpha_1, \dots, \alpha_h$, there exist an algebraic number field $K = \mathbb{Q}(\theta)$ which contains these numbers.

If α is an element of K , we define its size $|\overline{\alpha}|$ by the inequality

$$|\overline{\alpha}| = \max |r(\theta_i)|, \quad \alpha = r(\theta),$$

This definition agrees with the meaning of $|\overline{\alpha}|$ in Section 2.3. We now obviously have

$$|\overline{\alpha + \beta}| \leq |\overline{\alpha}| + |\overline{\beta}|, \quad |\overline{\alpha\beta}| \leq |\overline{\alpha}||\overline{\beta}|$$

The set of algebraic integers contained in the field of K forms a ring, which we shall call as \mathbb{Z}_K . We shall always have $\mathbb{Z} \subset \mathbb{Z}_K$.

Definition The *norm* of an element $\alpha \in K$ in the field K is defined by the product of all the conjugates of α in K . We denote the norm by $N(\alpha)$:

$$N(\alpha) = r(\theta_1) \dots r(\theta_h), \quad \alpha = r(\theta)$$

We shall now list and proof some of the properties of the Norm:

1. $N(\alpha) \in \mathbb{Q}$ for all $\alpha \in K$
2. $N(\alpha) \in \mathbb{Z}$ for all $\alpha \in \mathbb{Z}_K$
3. $N(\alpha) = 0$ if and onli if $\alpha = 0$,
4. $N(\alpha) = \alpha^h$ if $\alpha \in \mathbb{Q}$;
5. $N(\alpha \beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in K$;
6. $N(a \alpha) = a^h N(\alpha)$ for $a \in \mathbb{Q}$ and $\alpha \in K$.

Here, we will be giving the proofs of two of the above properties.

Claim:(3) $N(\alpha) = 0$ if and only if $\alpha = 0$

Proof. Suppose that $N(\alpha) = 0$,

$$\begin{aligned} N(\alpha) &= r(\theta_1) \dots r(\theta_h) = 0 \\ \Rightarrow r(\theta_i) &= 0, .i \in \{1, 2, \dots, h\} \end{aligned}$$

Let the minimal polynomial of θ be

$$\begin{aligned} p(x) &= (x - \theta_1) \dots (x - \theta_h), \\ p(\theta_i) &= 0 \end{aligned}$$

Recall that $p(x)$ is irreducible. So

$p(x)$ is also the min polynomial of θ_i . Hence $\{1, \theta_i, \dots, \theta_i^{h-1}\}$ is still linear independent

Hence

$$\begin{aligned} \alpha &= r(\theta) = a_0 + \dots + a_{h-1}\theta^{h-1}; \\ \Rightarrow r(\theta_i) &= a_0 + \dots + a_{h-1}\theta_i^{h-1} = 0. \end{aligned}$$

By linear independence property, we get

$$\begin{aligned} a_0 &= a_1 = \dots = a_{h-1} = 0; \\ \text{so } \alpha &= 0. \end{aligned}$$

Claim (5) : $N(\alpha\beta) = N(\alpha)N(\beta)$.

Proof: Let $\beta = s(\theta)$

Then

$$\begin{aligned} N(\alpha\beta) &= N(r(\theta)(s(\theta))) \\ &= [r(\theta_1)s(\theta_1)] \dots [r(\theta_h)s(\theta_h)] \\ &= [r(\theta_1) \dots r(\theta_h)][s(\theta_1) \dots s(\theta_h)] \\ &= N(\alpha)N(\beta). \end{aligned}$$

In what follows, when we are considering only a single $\alpha \in \mathbb{A}$, we shall always use $N(\alpha)$ to denote its norm in the algebraic number field $\mathbb{Q}(\alpha)$ generated by α .

We will now prove a lemma which is based on the symmetric polynomial theorem. Let V be a commutative ring with unit. A polynomial $F(\alpha_1, \dots, \alpha_n) \in V[\alpha_1, \dots, \alpha_n]$ is called a symmetric polynomial in $\alpha_1, \dots, \alpha_n$ if it does not change when $\alpha_1, \dots, \alpha_n$ are subjected to any permutation. We denote the following:

$$\begin{aligned}\sigma_1 &= \alpha_1 + \dots + \alpha_n, \\ \sigma_2 &= \alpha_1\alpha_2 + \dots + \alpha_{n-1}\alpha_n, \\ &\dots\dots\dots \\ \sigma_n &= \alpha_1 \dots \alpha_n,\end{aligned}$$

The above are called the elementary symmetric polynomials in $\alpha_1, \dots, \alpha_n$; they are (up to sign) the coefficients of the polynomial $(x-\alpha_1)\cdots(x-\alpha_n)$.

Exercise:

$$(x_1)^2 + (x_2)^2 + \dots + (x_n)^2 = \sum x_i^2$$

where

$$\begin{aligned}x_1 &= \alpha_1, \\ x_2 &= \alpha_2, \\ &\dots\dots\dots \\ x_n &= \alpha_n.\end{aligned}$$

Clearly, this summation is symmetric and can be expressed in terms of all the σ_i above

$$\begin{aligned}(x_1)^2 + (x_2)^2 + \dots + (x_n)^2 &= (x_1 + \dots + x_n)^2 - 2 \sum x_i x_j \\ &= \sigma_1^2 - 2\sigma_2 = H(\sigma_1, \sigma_2),\end{aligned}$$

where $H(Y_1, \dots, Y_n) = Y_1^2 - 2Y_2$.

The symmetric polynomial theorem states that:

any symmetric polynomial $F(\alpha_1, \dots, \alpha_n) \in V[\alpha_1, \dots, \alpha_n]$ can be uniquely expressed in the form of $F(\alpha_1, \dots, \alpha_n) = H[\sigma_1, \dots, \sigma_n]$, where $H[\sigma_1, \dots, \sigma_n] \in V[\sigma_1, \dots, \sigma_n]$.

Lemma 1. Suppose that $\alpha \in A$, $\deg(\alpha) = n$ and $\alpha = \alpha_1, \dots, \alpha_n$ are conjugates of α . Further suppose that

$$F = F(x_1, \dots, x_k; \alpha_1, \dots, \alpha_n) \in \mathbb{Q}(x_1, \dots, x_k; \alpha_1, \dots, \alpha_n), k \geq 0,$$

and that, as a polynomial in $\alpha_1, \dots, \alpha_n$ with coefficients in $\mathbb{Q}[x_1, \dots, x_k]$, F is a symmetric polynomial in $\alpha_1, \dots, \alpha_n$. Then

$$F = F(x_1, \dots, x_k) \in \mathbb{Q}[x_1, \dots, x_k];$$

and $F \in \mathbb{Q}$ in the case of $k = 0$. In addition, if we also have $\alpha \in \mathbb{Z}_A$ and

$$F = F(x_1, \dots, x_k; \alpha_1, \dots, \alpha_n) \in \mathbb{Z}(x_1, \dots, x_k; \alpha_1, \dots, \alpha_n), k \geq 0,$$

then it follows that

$$F = F(x_1, \dots, x_k) \in \mathbb{Z}[x_1, \dots, x_k],$$

and $F \in \mathbb{Z}$ in the case $k = 0$ (\Rightarrow there are no variables in the equation)

Proof. We shall consider F as a polynomial in $\alpha_1, \dots, \alpha_n$ with coefficients in $\mathbb{Q}[x_1 \dots x_k]$. Since F is a symmetric polynomial in $\alpha_1, \dots, \alpha_n$, it follows by the symmetric polynomial Theorem that it can be written as a polynomial in the elementary symmetric polynomials, $\sigma_1, \dots, \sigma_n$ with coefficients in $\mathbb{Q}[x_1 \dots x_k]$. But the elementary symmetric polynomials are equal up to sign to the coefficients of the minimal polynomial of the algebraic number α , and so they are elements of \mathbb{Q} . Hence, $F = F(x_1, \dots, x_k) \in \mathbb{Q}[x_1, \dots, x_k]$ and $F \in \mathbb{Q}$ on the case when $k = 0$.

We shall now proceed to Lemma 2:

Lemma 2. If

$$f(x) = b_n x^n + \dots + b_1 x + b_0, f(x) \in \mathbb{C}[x], \quad b_n b_0 \neq 0, \quad H_{f(x)} = H,$$

and $f(\alpha) = 0$, then the following inequalities hold:

$$\frac{|b_0|}{H + |b_0|} < |\alpha| < \frac{H + |b_n|}{b_n}, \quad |b_n \alpha| < 2H.$$

Proof. We shall prove the upper bound, then the lower bound.

For the upper bound: if $|\alpha| \leq 1$, then the upper bound obviously holds. Now suppose $|\alpha| > 1$.

Then recall that $f(\alpha) = 0$

$$\begin{aligned} |b_n||\alpha|^n &= |b_{n-1}\alpha^{n-1} + \dots + b_0| \leq H(|\alpha|^{n-1} + \dots + |\alpha| + 1) \\ &= H \frac{|\alpha|^n - 1}{|\alpha| - 1} < H \frac{|\alpha|^n}{|\alpha| - 1} \end{aligned}$$

from which we will obtain

$$|\alpha| = \frac{H + |b_n|}{|b_n|}$$

Thus, the lemma is proven for the upper bound.

Now, for the lower bound:

Define

$$\begin{aligned} g(x) &= x^n f(x^{-1}) \\ &= b_n + b_{n-1}x + \dots + b_0x^n \end{aligned}$$

Now, we will have

$$\begin{aligned} |b_0| \left| \frac{1}{\alpha} \right|^n &= |b_n + \dots + b_1| \left| \frac{1}{\alpha} \right|^{n-1} \\ &= H \left| 1 + \frac{1}{|\alpha|} + \dots \right| \\ &= H \frac{1/|\alpha|^n - 1}{1/|\alpha| - 1} = H \frac{1 - |\alpha|^n}{1 - |\alpha|} \frac{1}{|\alpha|^{n-1}} \\ &< \frac{H}{1 - |\alpha|} \frac{1}{|\alpha|^{n-1}} \end{aligned}$$

so that $|b_0|(1 - |\alpha|) < H|\alpha|$

$$\Rightarrow |\alpha| > \frac{|b_0|}{H + |b_0|}$$

Thus, the lemma is proven for the lower bound.

Corollary. If α is a non-zero algebraic integer of height h , then its size satisfies the bound of

$$\frac{1}{2h} < \overline{|\alpha|} < h + 1 \leq 2h. \quad (2.43)$$

Proof. Clearly,

$$\overline{|\alpha|} = \max |\alpha| < \frac{H}{|b_n|} + 1 < h + 1 < 2h \quad \square$$

Theorem 8. Suppose that $\alpha \in \mathbb{A}$, $\deg \alpha \leq n$, $H_\alpha \leq h$, and

$$P = P(x) \in \mathbb{Z}[x], \quad \deg(P) \leq k, H_p \leq H.$$

Then either $P(\alpha) = 0$ or else the following inequality holds:

$$|P(\alpha)| \geq \frac{c^k}{H^{n-1}}, \quad c = \frac{1}{3^{n-1}h^n}. \quad (2.44)$$

Proof. Without loss in generality, we may assume that $\deg \alpha = n$, $H_\alpha = h$, $\deg P = k$, $H_p = H$. Suppose that $P(\alpha) \neq 0$. Then $H \geq 1$. We let b_n denote the leading coefficient in the primitive and irreducible polynomial in $\mathbb{Z}[x]$ having α as a root, where $b_n > 0$. Then $\epsilon = b_n^k P(\alpha) \in \mathbb{Z}_A$. Of $n = 1$, then

$$|\epsilon| = b_n^k |P(\alpha)| \geq 1, \quad |P(\alpha)| \geq \frac{1}{b_n^k} \geq \frac{1}{h^k},$$

which proves (2.44) in this case. Now suppose that $n > 1$. We have

$$|N(\epsilon)| = b_n^k |P(\alpha)| \prod_{i=2}^n b_n^k |P(\alpha_i)| \geq 1 \quad (2.45)$$

Observing the above equality closer, we find that by using lemma 2,

$$\begin{aligned} b_n^k |P(\alpha_i)| \leq b_n^k H(1 + |\alpha_i| + \cdots + |\alpha_i|^k) &< b_n^k H(1 + |\alpha_i|)^k \\ &\leq b_n^k H \left(1 + \frac{H + |b_n|}{|b_n|}\right)^k \\ &\leq H(b_n + h + b_n)^k \\ &\leq H(2b_n + h)^k \\ &\leq H(2h + h)^k \\ &\leq H(3h)^k. \end{aligned} \quad (2.46)$$

From (2.45) and (2.46), we have

$$|P(\alpha)| > \frac{1}{b_n^k H^{n-1} (3h)^{k(n-1)}} \geq \frac{1}{H^{n-1} (3^{n-1} h^n)^k}.$$

As this gives us the required form, we have completed the proof for Theorem 8.

Theorem 9 If α is an algebraic number of degree n , $n \geq 1$, then there exist a constant $c = c(\alpha) > 0$, such that the following inequality holds for any algebraic number θ of degree k , $k \geq 1$, and height H for which $\theta \neq \alpha$:

$$|\alpha - \theta| > \frac{c^k}{H^n}. \quad (2.47)$$

Proof. Suppose that θ is not equal to α and is a root of the irreducible polynomial

$$P(x) = a_k x^k + \cdots + a_0, \quad P(x) \in \mathbb{Z}[x], \quad a_k > 0. \quad (2.48)$$

There are 3 possibilities:

Case 1. $k = 1$. In this case, the theorem simply follows from Liouville's Theorem.

Case 2. $k > 1$ and $P(\alpha) = 0$. Then $k = n$ and θ is a conjugate of α . Let $\delta = \delta(\alpha)$ be any constant such that

$$0 < \delta < \min |\alpha_i - \alpha_j|.$$

Then,

$$|\alpha - \theta| > \delta \geq \frac{\sqrt[n]{\delta^k}}{H^n} = \frac{c^k}{H^n}. \quad (2.49)$$

The inequality (2.49) shows that the theorem is true in this case.

Case 3. $k > 1$ and $P(\alpha) \neq 0$. If $|\alpha - \theta| \geq 1$, then

$$|\alpha - \theta| \geq \frac{1}{H^n} \quad (2.50)$$

On the other hand, if $|\alpha - \theta| < 1$, then

$$|\theta| < |\alpha| + 1. \quad (2.51)$$

We write

$$|\alpha - \theta| = \frac{|P(\alpha)|}{|P_1(\alpha)|} \quad (2.52)$$

and we bound $|P_1(\alpha)|$ from above. We have

$$\begin{aligned} P_1(x) &= \frac{P(x)}{x - \theta} = \frac{P(x) - P(\theta)}{x - \theta} \\ &= a_k \frac{x^k - \theta^k}{x - \theta} + \cdots + a_1 \frac{x - \theta}{x - \theta} \\ &= a_k(x^{k-1} + \theta x^{k-2} + \cdots + \theta^{k-1}) + a_{k-1}(x^{k-2} + \theta^{k-3} + \cdots + \theta^{k-2}) + \cdots + a_1 \\ &= \sum_{l=0}^{k-1} g_l(\theta) x^l. \end{aligned} \quad (2.53)$$

Where

$$g_l(\theta) = a_{l+1} + \cdots + a_k \theta^{k-l-1}, \quad l = 0, 1, \dots, k-1 \quad (2.54)$$

From the above few equations (namely (2.54) and (2.51)), we find that,

$$\begin{aligned} |g_l(\theta)| &< H(1 + |\theta| + \cdots + |\theta^k|) \leq H(1 + |\theta|)^k \\ &< (|\alpha| + 2)^k H, \quad l = 0, 1, \dots, k-1 \end{aligned} \quad (2.55)$$

and from (2.53) and (2.55), if we use the previous corollary to Lemma 2 (see (2.44)), we find that

$$\begin{aligned} |P_1(\alpha)| &< (|\alpha| + 2)^k H(1 + |\alpha| + \cdots + |\alpha|^k) \\ &\leq (|\alpha| + 2)^k (|\alpha| + 1)^k H < ((h+2)(h+3))^k H \end{aligned} \quad (2.56)$$

By theorem 8, there exists a constant $c_0 = c_0(\alpha)$, $0 < c_0 \leq 1$ such that

$$|P(\alpha)| > \frac{c_0^k}{H^{n-1}}, \quad c_0 = \frac{1}{3^{n-1} h^n} \quad (2.57)$$

substitute (2.56) and (2.57) into (2.52), we will get the bound of

$$|\alpha - \theta| > \frac{c^k}{H^n}, \quad c = \frac{c_0}{(h+2)(h+3)} < 1. \quad (2.58)$$

Now, the desired inequality (2.49) will be given by combining (2.50) and (2.58).

The Theorem is thus proven.

Theorem 10. Suppose that α is a complex number such that for any $v \in \mathbb{R}$, $v > 0$ the inequality

$$|\alpha - \theta| < \frac{1}{H_\theta^v}$$

has an infinite set of solutions in algebraic numbers θ of degree $\deg \theta \leq k$ and height H_θ . Then α is transcendental.

2.5 Further refinements and generalizations of Liouville's Theorem

Liouville's Theorem gives an upper bound for the order of approximation of an algebraic number by rational numbers. Hence, after the theorem was published, the question naturally arose: *Given an algebraic number α of degree n , $n \geq 3$, find a constant $v=v(\alpha)$ such that the inequality*

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^{v+\epsilon}} \quad (2.59)$$

has only infinitely many integer solutions $p \in \mathbb{Z}$, $q \in \mathbb{N}$ if $\epsilon > 0$ and infinitely many solution if $\epsilon < 0$. From Liouville's Theorem, it follows that $v(\alpha) \leq n$ for any algebraic number α of degree n . In 1909, Thue published a method for obtaining results on the approximation of algebraic numbers. He used this method to lower significantly the bound for $v(\alpha)$. He show that $v(\alpha) \leq n/2 + 1$ for any algebraic number α of degree $n \geq 2$. In 1908, he obtained a similar result, but only for numbers of the form $\sqrt[n]{a/b}$, where $a, b \in \mathbb{N}$

Thue's Theorem If α is an algebraic number of degree n , $n \geq 2$, then if ϵ is any positive number, then the inequality

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^{n/2+1+\epsilon}} \quad (2.60)$$

has only finitely many solutions $p \in \mathbb{Z}$, $q \in \mathbb{N}$

In 1955, K.F. Roth came up with another Theorem

Roth's Theorem if α is an algebraic number of degree n , $n \geq 2$, and if ϵ is any positive number, then the inequality,

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^{2+\epsilon}}$$

We also come up with a version which is known as

Thue-Siegel-Roth Theorem . Let α be an algebraic number with $\alpha \notin \mathbb{Q}$. Let $\epsilon > 0$. Then there are at most finitely many pairs of integers (a, b) with $b > 0$ such that

$$\left| \alpha - \frac{a}{b} \right| < \frac{1}{b^{2+\epsilon}}$$

Chapter 3

Arithmetic Properties of the values of the exponential function at algebraic points

3.1 Transcendence of e

Theorem 1. *The number e is irrational*

Proof. e is given by the series

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} \quad (3.1)$$

We denote

$$A_n = n! \sum_{k=0}^n \frac{1}{k!} \in \mathbb{N}, \quad n = 1, 2, \dots; \quad (3.2)$$

$$a_n = n! \sum_{k=n+1}^{\infty} \frac{1}{k!} > 0, \quad n = 1, 2, \dots; \quad (3.3)$$

We have

$$\begin{aligned} a_n &= \frac{1}{n+1} \left(1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \dots \right) < \\ &< \frac{1}{n+1} \left(1 + \frac{1}{2} + \frac{1}{2^2} + \dots \right) = \frac{2}{n+1}, \quad n = 1, 2, \dots \end{aligned} \quad (3.4)$$

Combining (3.3) and (3.4), we obtain:

$$0 < a_n < \frac{2}{n+1} \leq 1, \quad n = 1, 2, \dots \quad (3.5)$$

Using (3.1)-(3.3) and (3.5), we see that

$$n!e = (A_n + a_n) \notin \mathbb{N}, \quad n = 1, 2, \dots \quad (3.6)$$

Proving by Contradiction.

Now suppose that e is a rational number: $e = p/q$ with $p, q \in \mathbb{N}$.

Setting $n=q$, we have $ne \in \mathbb{N}$, which contradicts (3.6).

Thus, the theorem is proven.

Theorem 2 e is not a quadratic irrationality

Proof Suppose the contrary, that is

$$ae^2 + be + c = 0, \quad a, b, c \in \mathbb{Z}$$

where a, b, c are not all zero. By theorem1, we have $a \neq 0$ and $c \neq 0$. We

now suppose that $a > 0$. We have

$$ae + b + ce^{-1} = 0, \quad a > 0, c \neq 0 \quad (3.7)$$

The number $1/e$ is given by the series

$$\frac{1}{e} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \quad (3.8)$$

we denote

$$B_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}, \quad B_n \in \mathbb{Z}, \quad n = 1, 2, \dots \quad (3.9)$$

and

$$\begin{aligned} b_n &= n! \sum_{k=m+1}^{\infty} \frac{(-1)^{k-n-1}}{k!} \\ &= \frac{1}{n+1} - \frac{1}{(n+1)(n+2)} + \frac{1}{(n+1)(n+2)(n+3)} - \dots \end{aligned} \quad (3.10)$$

Note that b_n is given by an alternating series whose terms decrease monotonically in absolute value. Hence,

$$0 < \frac{1}{n+1} - \frac{1}{(n+1)(n+2)} < b_n < \frac{1}{n+1}. \quad (3.11)$$

From (3.1), (3.7) and (3.8), along with the definitions (3.2),(3.3),(3.9) and (3.10), we get

$$\begin{aligned} n!(ae + b + ce^{-1}) &= (nA_n + bn! + cB_n) + (aa_n + (-1)^{n+1}cb_n) = \\ &= C_n + c_n = 0. \end{aligned} \quad (3.12)$$

By (3.2) and (3.9), we have

$$C_n = (aA_n + bn! + cB_n) \in \mathbb{Z}, \quad n = 1, 2, \dots \quad (3.13)$$

We choose an n satisfying the two inequalities

$$n \geq 2a + |c|, \quad (-1)^{n+1}c > 0. \quad (3.14)$$

Since $a > 0$, we now recall that

$$0 < c_n = aa_n + (-1)^{n+1}cb_n < \frac{2a + |c|}{n+1}.$$

Recall that

$$\begin{aligned} 0 < a_n < \frac{2}{n+1}, \quad , 0 < b_n < \frac{1}{n+1}; \\ 0 < c_n < \frac{2a}{n+1} + \frac{|c|}{n+1} \leq \frac{n}{n+1} < 1. \end{aligned} \quad (3.15)$$

But the condition (3.13) and the inequality (3.15) contradicts the equation (3.12).

As

$$C_n + c_n \neq 0.$$

ie, we have a contradiction. Hence, we have proven the $ae^2 + be + c \neq 0$.

Lemma I. If $g(x) \in \mathbb{Z}[x]$ for any $k \in \mathbb{N}$ all of the coefficients of the k -th derivative $g^k(x)$ are divisible by $k!$.

Proof. Since differentiation is a linear operation, it suffices to prove that the lemma for the polynomial x^k , $s > 0$. But the k -th derivative of x^k is zero if $k > s$, and if $1 \leq k \leq s$, then it's equal to $k! \binom{s}{k} x^{s-k}$, in which the binomial coefficient $\binom{s}{k}$ is an integer. The lemma is thus proven!

Lemma 2. Let $f(x)$ be a polynomial of degree v with real coefficients, and set

$$F(x) = f(x) + f^1(x) + \cdots + f^v(x). \quad (3.16)$$

Then

$$e^x \int_0^x f(t)e^{-t} dt = F(0)e^x - F(x), \quad (3.17)$$

Where x is a real or complex number.

Proof. Integrating by parts, we obtain the relation

$$\begin{aligned} \int_0^x f(t)e^{-t} dt &= -f(t)e^{-t} + \int_0^x f^1(t)e^{-t} dt \\ &= f(0) - f(x)e^{-x} + \int_0^x f^1(t)e^{-t} dt \end{aligned} \quad (3.18)$$

We continue this process for $(v+1)$ times

We will get

$$\begin{aligned} \int_0^x f(t)e^{-t} dt &= f(0) + -e^{-x}f(x) + f^1(0) - e^{-x}f^1(x) + \cdots \\ &= F(0) - e^{-x}F(x) \end{aligned}$$

Then

$$e^x \int_0^x f(t)e^{-t} dt = F(0)e^x - F(x).$$

Equation (3.17) is known as the Hermite identity.

Theorem 3. (Hermite). e is *transcendental*. *Proof* We suppose the contrary, ie, that e is an algebraic number of degree m .

Then

$$a_m e^m + \cdots + a_1 e + a_0 = 0, \quad a_0 \neq 0, \quad a_k \in \mathbb{Z}, (k = 0, 1, \dots, m). \quad (3.19)$$

If we set $x = k$ in Hermite's identity in (3.17), where $k = 0, 1, \dots, m$, we obtain:

$$e^k \int_0^k f(t)e^{-t} dt = F(0)e^k - F(k), \quad (k = 0, 1, \dots, m). \quad (3.20)$$

We multiply both sides of (3.20) by a_k , and then add the resulting equations for $k = 0, 1, \dots, m$. Assuming $a_m e^m + \dots + a_1 e + a_0 = 0$, we find that

$$-\sum_{k=0}^m a_k F(k) = \sum_{k=0}^m a_k e^k \int_0^k f(t) e^{-t} dt. \quad (3.21)$$

Since the equation above holds for any polynomial $f(x)$ with real coefficients, we can simply let

$$f(x) = \frac{1}{(n-1)!} x^{n-1} (x-1)^n \dots (x-m)^n, \quad (3.22)$$

Where n (a large natural number) which can be chosen by some conditions to be set later.

Note: In the above equation for $f(x)$, $f(x)$ has zero as a root of multiplicity of order $n-1$, and $1, \dots, m$ as a root of multiplicity of order n .

Brief idea of proof from this point on: We show that the left hand side of (3.21) will be some non-zero rational integer, while the right side will have absolute value less than 1. This contradiction will prove the theorem.

We will now write out the remaining of the proof in detail.

Differentiating the function f :

$$f^l(0) = 0, \quad l = 0, 1, \dots, n-2 \quad (3.23)$$

$$f^{n-1}(0) = (-1)^{mn} (m!)^n, \quad (3.24)$$

$$f^l(k) = 0 \quad l = 0, 1, \dots, n-1; k = 1, \dots, m. \quad (3.25)$$

Using lemma 1, the l -th derivative of $x^{n-1} x - 1^n \dots x - m^n$ has coefficients which are divisible by $l!$. This implies that for $l \geq n$ the coefficients of $f^l(x)$ are divisible by n . Hence, from (3.16), (3.23) and (3.24), we find that,

$$F(0) = f(0) + f^1(0) + \dots + f^v(0),$$

Since $f(0) = 0, f'(0) = 0, \dots, f^l(0) = 0$ for l from 1 to $(n-2)$ we find that

$$F(0) = \sum_{l=n-1}^{(m+n)n-1} f^l(0) = -1^{mn} m!^n + nA, \quad A \in \mathbb{Z} \quad (3.26)$$

$v = \deg \text{ of } f(x) = mn + n - 1 = (m + n)n - 1$

Proof of (3.26.) Remember that

$$f(x) = \frac{1}{(n-1)!} x^{n-1} (x-1)^n \cdots (x-m)^n,$$

$$f^n(0) = \frac{1}{(n-1)!} n! C = nC, \text{ where } C = \text{constant}$$

$$\text{and } -1^{mn} m!^n = f^{n-1}(0)$$

And we also find that

$$F(k) = \sum_{l=n-1}^{(m+n)n-1} f^l(k) = nB_k, \quad B_k \in \mathbb{Z} \quad (3.27)$$

Now we define n as any integer satisfying the conditions

$$\text{GCD}(n, m!) = 1, \quad n > |a_0| \quad (3.28)$$

We then come up with some important observations:

Conditions include:

Condition 1: $a_0 F(0)$ is not divisible by n

Proof Given

$$F(0) = \sum_{l=n-1}^{(m+n)n-1} f^l(0) = (-1)^{mn} m!^n + nA$$

when we multiply a_0 to both sides, we get $a_0 F(0) = a_0 (-1)^{mn} m!^n + a_0 nA$

If $a_0 F(0) \equiv 0 \pmod{n}$

$$\Rightarrow a_0 [(-1)^{mn} m!^n + nA] \equiv 0 \pmod{n}$$

$$\Rightarrow n \mid a_0 - 1^{mn} (m!)^n$$

$$\Rightarrow n \mid a_0 \text{ which contradicts (3.28)}$$

Hence, $a_0 F(0)$ is not divisible by n is proven.

Condition 2: The left hand side of (3.21) is a non-zero integer.

Proof We observe from (3.27) that LHS of (3.21), $-\sum_0^m a_k F(k) \in \mathbb{Z}$ and prove

instead that it is non-zero.

Prove by contradiction:

Assume left-hand side of (3.21) is zero,

$$\begin{aligned} &= -\sum_{k=0}^m a_k F(k) \\ &= -a_0 F(0) - \sum_{k=0}^m a_k F(k) \\ &= 0 \end{aligned}$$

then

$$\begin{aligned} &\Rightarrow -a_0 F(0) - n(B_1 + B_2 + \cdots + B_m) = 0 \\ &\Rightarrow -a_0 F(0) = n(B_1 + B_2 + \cdots + B_m) \\ &\Rightarrow n \mid a_0 F(0) \text{ which is once again a contradiction,} \end{aligned}$$

Hence, $-\sum_{k=0}^m a_k F(k)$ is indeed a non-zero integer.

In particular, we get from the above 2 conditions that

$$\left| \left(\sum_{k=0}^m a_k F(k) \right) \right| \geq 1 \quad (3.29)$$

Next, we shall find an upper bound for the right side of (3.21). On the interval $0 \leq x \leq m$, each of the factors $x - k$ in (3.22), $0 \leq x \leq m$, is bounded by m . Thus,

$$|(f(x))| \leq \frac{m^{(m+1)n-1}}{(n-1)!}, \quad 0 \leq x \leq m$$

and since $f(t)$ has a upper bound of $\frac{m^{(m+1)n-1}}{(n-1)!}$,

$$\begin{aligned} \left| \sum_{k=0}^m a_k \int_0^k f(t) e^{k-t} dt \right| &\leq \frac{m^{(m+1)n-1}}{(n-1)!} \sum_{k=0}^m |a_k| \int_0^k f(t) e^{k-t} dt \\ &< \frac{m^{(m+1)n}}{(n-1)!} e^m \sum_{k=0}^m |a_k| \\ &= C_0 \frac{C^n}{(n-1)!}, \end{aligned} \quad (3.30)$$

Where C_0 and C does not depend on n . From (3.21) and (3.29)-(3.30), we obtain the inequality

$$1 \leq \left| \left(\sum_{k=0}^m a_k F(k) \right) \right| < C_0 \frac{C^n}{(n-1)!},$$

If we compare the Left-hand side and Right-hand side of (3.21), we get

$$\text{LHS} = |(\sum_{k=0}^m a_k F(k))| \geq 1$$

$$\text{RHS} = |\sum_{k=0}^m a_k \int_0^k f(t) e^{k-t} dt| \leq C_0 \frac{C^n}{(n-1)!} \rightarrow 0 \text{ as } n \rightarrow \infty$$

we have a contradiction,

Hence, our original assumption that e is algebraic is proven wrong.

Hence, e MUST BE *transcendental*.

Recall that

$$f(x) = \frac{1}{(n-1)!} x^{n-1} (x-1)^n \cdots (x-m)^n, \tag{3.31}$$

where $(x)^{n-1}$, $(x-1)^n$, \cdots are all bounded by m^n as on the interval $0 \leq x \leq m$, each of the factors $(x-k)$ are bounded by m . Hence, by estimation, we get the right hand side of (3.20) as

$$|f(x)| \leq \frac{m^{(m+1)n}}{(n-1)!} \tag{3.32}$$

ie, RHS tends to zero as n tends to infinity.

Therefore, we will get

$$\Rightarrow F(0)e^x = F(x) \text{ if } n \rightarrow \infty$$

$$\Rightarrow e^x = \frac{F(x)}{F(0)}$$

$$\Rightarrow e^k = \frac{F(k)}{F(0)}$$

Hence for every n , the fractions $\frac{F(k)}{F(0)}$ are simultaneous approximations to the value e^k , $k=1, \dots, m$.

Thus the above proof that e is transcendental is based on a construction (using Hermite's identity) of a sequence of simultaneous approximations of powers of e .

3.2 Lindemann's theorem

Our next result is due to Lindemann.

Theorem 1. Let $\alpha_1, \dots, \alpha_n$ be distinct algebraic numbers, and let β_1, \dots, β_n be non-zero algebraic numbers. Then

$$\beta_1 e^{\alpha_1} + \dots + \beta_n e^{\alpha_n} \neq 0. \quad (3.33)$$

The number e^{α_j} above may be multi-valued. The theorem is true for any values of e^{α_j} .

We shall exclude the proof of Lindemann's theorem, and instead, look at the results that we can deduce from the above theorem:

Corollary The following numbers are transcendental:

- (i) π
- (ii) e^α if α is a non-zero algebraic number.
- (iii) $\sin(\alpha)$, $\cos(\alpha)$, and $\tan(\alpha)$ if α is a non-zero algebraic number.
- (iv) $\log(\alpha)$ if α is an algebraic number different from 0 and 1.

Claim: (ii) If α is a non-zero algebraic number, then e^α is transcendental.

Proof:

Case 1) If $\alpha = 1$, then $e^\alpha = e$ is transcendental (as proven before).

Case 2) If $\alpha \neq 1$, then by Lindemann's Theorem, we have $e^0 = 1$ and e^α are linearly independent over algebraic numbers.

But 1 is a non-zero algebraic number, hence this will imply that e^α is transcendental.

Claim: (iii) $\sin(\alpha)$, $\cos(\alpha)$, and $\tan(\alpha)$ are transcendental if α is a non-zero algebraic number.

Proof:

We know that $\sin(\alpha) = \frac{1}{2}(e^{i\alpha} + e^{-i\alpha})$ and $\cos(\alpha) = \frac{1}{2}(e^{i\alpha} - e^{-i\alpha})$. Hence, $\sin(\alpha)$ and $\cos(\alpha)$ are both linear independent over algebraic numbers.

So not both of them are algebraic numbers, ie at least one of them is transcendental.

But remembering that $\sin^2(\alpha) + \cos^2(\alpha) = 1$,

hence both $\sin(\alpha)$ and $\cos(\alpha)$ are transcendental.

Claim: (iv) $\log(\alpha)$ is transcendental if α is an algebraic number different from 0 and 1.

Proof:

Assuming α is an algebraic number.

If $\log(\alpha)$ is algebraic, then $\alpha = e^{\log \alpha}$ is transcendental by Claim (ii), which is a contradiction.

Hence $\log(\alpha)$ must be transcendental.

3.3 The Gelfond-Schneider Theorem and Some Related Results

We shall start with stating some results without proofs.

In the 1900, David Hilbert posed a general question, which included determining whether $2^{\sqrt{2}}$ is transcendental and whether e^π is transcendental. The problem was resolved independently by Gelfond and Schneider in 1934. Their result is the following

Gelfond-Schneider Theorem If α and β are algebraic numbers with $\alpha \neq 0$, $\alpha \neq 1$, and $\beta \notin \mathbb{Q}$, then α^β is transcendental.

Observe that the theorem asserts that any value of α^β is transcendental under the conditions above. It is clear that $2^{\sqrt{2}}$ (The Gelfond-Schneider Constant) is transcendental from this result, and since e^π is a value of i^{-2i} , the transcendence of e^π also follows from this result. We note that the following are equivalent forms of this result:

(i) If l and β are complex numbers with l not equal to 0 and β not inside \mathbb{Q} , then at least one of the three numbers e^l , β and $e^{l\beta}$ is transcendental.

(ii) If α and β are non-zero algebraic numbers with $\log(\alpha)$ and $\log(\beta)$ linearly independent over the rationals, then $\log(\alpha)$ and $\log(\beta)$ are linearly independent over the algebraic

numbers.

We shall observe that (ii) is the same as saying that if α and β are non-zero algebraic numbers with β not equal to 1 and $\log(\alpha)/\log(\beta) \notin \mathbb{Q}$, then $\log(\alpha)/\log(\beta)$ is transcendental.

Proofs of equivalences.

Gelfond-Schneider Theorem \Rightarrow (i):

We can just take α as e^l . Then clearly, α is not 0 or 1. Then Gelfond theorem states that if $\alpha(=e^l)$ and β are algebraic, then $\alpha^\beta = e^{l\beta}$ is transcendental, which implies (i).

(i) \Rightarrow (ii):

We first observe that the condition $\log(\alpha)$ and $\log(\beta)$ are linearly independent over the rational implies that both α and β are not 1. Also, we get that $\log(\alpha)/\log(\beta) \notin \mathbb{Q}$. So we just let $l = \log(\beta)$ and $\beta' = \log(\alpha)/\log(\beta)$.

Then (i) implies that one of the following numbers are transcendental: $e^l(= \beta)$, β' or $e^{l\beta'}(= \alpha)$

But since α and β are given to be algebraic, the number that is transcendental is thus β' , ie, $\beta' = \log(\alpha)/\log(\beta) = p$, where p is any transcendental number.

This implies that $\log(\alpha)$ and $\log(\beta)$ are linearly independent over the algebraic numbers. Hence, (i) implies (ii).

(ii) \Rightarrow Gelfond-Schneider Theorem:

Take α and β as algebraic numbers. Assume that α^β algebraic,

We consider $\beta' = e^{\beta(\log\alpha)}$. Then $\log(\alpha)$ and $\log(\beta')(= \beta\log(\alpha))$ are linearly dependent over the algebraic numbers. Hence, by (ii), we get $\log(\alpha)$ and $\beta\log(\alpha)$ are linearly dependent over the rationals. This implies

$$a(\log\alpha) + b\beta(\log\alpha) = 0$$

$$a + b\beta = 0$$

$$\Rightarrow \beta = \frac{a}{b}$$

Hence, we will get $\beta \in \mathbb{Q}$. Hence, (ii) implies Gelfond-Schneider Theorem.

There are some results that are similar to (i). For example, Lang had proven that

Theorem. Suppose l_1, l_2 and l_3 are linearly independent over the rationals and that β_1 and β_2 are linearly independent over the rationals. Then at least one of the numbers $e^{l_i \beta_j}$ is transcendental.

In 1966, Baker established the following generalization of the Gelfond-Schneider Theorem:

Theorem. If $\alpha_1, \dots, \alpha_m$ are non-zero algebraic numbers with $\log \alpha_1, \dots, \log \alpha_m$ linearly independent over the rationals, then $\log \alpha_1, \dots, \alpha_m$ are linearly independent over the algebraic numbers.

Chapter 4

The computation of Transcendental Functions

4.1 Introduction

Transcendental functions can be computed in software by a variety of algorithms [3]. The algorithms that are most suitable for implementation on modern computer architectures usually comprise of three steps: *reduction*, *approximation*, and *reconstruction*.

These steps are best illustrated by an example. Consider the calculation of the exponential function e^x . One may first attempt an evaluation using the familiar Maclaurin series expansion:

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^k}{k!} + \dots \quad (4.1)$$

When x is small, computing a few terms of this series gives a reasonably good approximation to e^x up to, for example, IEEE double precision (which is approximately 17 significant decimal digits) [4]. However, when x is large, many more terms of the series are needed to satisfy the same accuracy requirements. Increasing the number of terms not only lengthens the calculation, but it also introduces more accumulated rounding errors that may degrade the accuracy of the answer.

To solve this problem, we express x as

$$x = N\left(\frac{\ln 2}{2^K}\right) + r. \quad N \in \mathbb{Z}, r \in \mathbb{R} \quad (4.2)$$

for some integer K chosen beforehand.

If we make $N\left(\frac{\ln 2}{2^K}\right)$ to be as close to x as possible, then $|r|$ never exceeds $\left(\frac{\ln 2}{2^{K+1}}\right)$.

The mathematical identity

$$e^x = e^{N\left(\frac{\ln 2}{2^K}\right)+r} = e^{\ln 2^{\frac{N}{2^K}}} e^r = 2^{\frac{N}{2^K}} e^r. \quad (4.3)$$

shows that the problem is transformed to that of calculating the $\exp(e)$ function at an argument of r whose magnitude is confined.

The transformation to r from x is called the *reduction step*;

the calculation of e^r , usually performed by computing an approximating polynomial, is called the *approximation step*;

and the composition of the final results based on e^r and the constant related to N and K is called the *reconstruction step*.

In a more traditional approach, K is chosen to be 0 and thus the approximation step requires a polynomial with accuracy good to IEEE double precision for the range of $|r| \leq \ln(2)/2$. This choice of K leads to reconstruction via multiplication of e^r by 2^N , which is easily implemented, for example, by scaling the exponent field of a floating-point number. One drawback of this approach of this approach is when $|r|$ is near $\ln(2)/2$, a large number of terms of the Maclaurin series expansion is still needed.

More recently, a framework known as *table-driven algorithm* suggests the use of $K > 1$. When for example $K = 5$, the argument r after the reduction step would satisfy $|r| \leq \ln(2)/64$. As a result, a much shorter polynomial can satisfy the same accuracy requirement. The tradeoff is a more complex reconstruction step, requiring the multiplication with a constant of the form

$$2^{\frac{N}{32}} = 2^M 2^{\frac{d}{32}}, d = 0, 1, \dots, 31, \quad (4.4)$$

where $N = 32M + d$.

This constant can be obtained rather easily, provided that all the 32 possible values of the 2nd factor are computed beforehand and stored in a table(hence the name 'table-driven'). This framework works well for modern machines not only because tables(even large ones) can be accomodated, but also because parallelism, such as the presence of pipelined arithmetic units, allows most of the extra work in the reconstruction step to be carried out while the approximating polynomial is being evaluated. This extra work includes, for example, the calculation of d , the indexing into and the fetching from the table of constants, and the multiplication to form $2^{N/32}$. Consequently, the performance gain due to a shorter polynomial is fully realised.

Example. Now we will illustrate what we have discussed so far by coming up with an example on how to calculate exponential functions,

To calculate e^{50}

Reduction Step Assuming that $K = 5$, We will get the argument r such that

$$|r| \leq \frac{\ln 2}{64} = 0.0108$$

We then choose a value for r such that it is close to its bound of $\ln 2/64$.

We thus choose $r = 0.01$.

Approximation Step Using Maclaurin Series, we find e^r , where r is assumed to be what we had chosen before;

$$\begin{aligned} e^r &= 1 + r + \frac{r^2}{2!} + \frac{r^3}{3!} \dots \\ &= 1.010050167 \end{aligned}$$

by calculating the first 5th terms of the Maclaurin Series.

Reconstruction Step

If $x = 50$, and $r = 0.01$ and recall that $x = N(\frac{\ln 2}{2^K}) + r$, then

$$N = \frac{2^5}{\ln 2}(50 - 0.01),$$

$$\simeq 2308. \tag{4.5}$$

Then

$$\begin{aligned} 2^{\frac{2308}{32}} &= 2^{72} \cdot 2^{\frac{4}{32}}, \\ &= 5.15 \times 10^{21} \end{aligned}$$

Therefore,

$$e^{50} = 2^{\frac{2308}{32}} \cdot e^r. \tag{4.6}$$

Hence, we get $e^{50} \simeq 5.201 \times 10^{21}$.

The above example illustrates clearly the advantages of calculating in this manner. If we had not used a transformation, we might have been required to evaluate the Maclaurin Series beyond the 5th element. Instead, in our calculations above, we only need to evaluate up to the 5th element for Maclaurin Series defining e^r to get a very accurate computation of the exponential function at $x = 50$. \square

We shall now conduct investigations into how the choices of K will affect the computations of the Exponential Functions.

Investigations of the effects of varying the values of K

We shall fix x as 5, and determine the values of e^5 for different K s.

A table containing the results(different values for e^5) are given below:

K	e^5
0	1.571
3	3.577
5	1.468×10^2
8	2.32×10^{17}
10	3.039×10^{69}

We shall provide the working that $e^5 \approx 1.468 \times 10^2$ for $K = 5$.

Working:

Reduction Step Assuming that $K = 5$, We will get the argument r such that

$$|r| \leq \frac{\ln 2}{64} = 0.0108$$

We then choose a value for r such that it is close to its bound of $\ln 2/64$.

We thus choose $r = 0.01$.

Approximation Step Using Maclaurin Series, we find e^r , where r is assumed to be what we had chosen before;

$$\begin{aligned}e^r &= 1 + r + \frac{r^2}{2!} + \frac{r^3}{3!} \cdots \\ &= 1.010050167\end{aligned}$$

by calculating the first 5th terms of the Maclaurin Series.

Reconstruction Step

If $x = 5$, and $r = 0.01$ and recall that $x = N(\frac{\ln 2}{2^K}) + r$, then

$$\begin{aligned}N &= \frac{2^5}{\ln 2}(5 - 0.01), \\ &\simeq 230.\end{aligned}$$

Then

$$\begin{aligned}2^{\frac{230}{32}} &= 2^7 \cdot 2^{\frac{6}{32}}, \\ &= 1.457 \times 10^2\end{aligned}$$

Therefore,

$$e^5 = 2^{\frac{230}{32}} \cdot e^r.$$

Hence, we get $e^5 \simeq 1.468 \times 10^2$. \square

Conclusion We can easily see from the table on the previous page that as K differs from its so-called threshold value of 5, the value that we get as e^5 also tends to become either too large or too small. Hence, we can easily deduce that the most suitable value that we can use for the K value would be 5. \square

Next, we shall look at the significance of the computational method used in this chapter. We shall show this in the form of a table,

Taking the value of K as 5

The Exponential Function	Method 1	Method 2	Actual Values
e^5	1.468x10 ²	65.375	1.483x10 ²
e^{10}	2.193x10 ⁴	644.33	2.200x10 ⁴
e^{20}	4.762x10 ⁸	8221	4.484x10 ⁸
e^{30}	1.0797x10 ¹³	38731	1.0653x10 ¹³

Note: Method 1 denotes the answer we get by applying the computational method used in this chapter,

Method 2 denotes using the Maclaurin series for each e^x .

Both methods uses only the first 5 terms in the Maclaurin's series to get to their respective answers.

The last column shows the actual value of each e^x , assuming $e = 2.718$.

Conclusion: It is not hard to notice that the values in Method 1 are closer to the actual value for each e^x than those gotten in Method 2. Hence, we can safely deduce that using this computational method to calculate the exponential functions is a more efficient (and accurate) way of computing. \square

Bibliography

- [1] Transcendental Number Theory:
<http://www.math.sc.edu/filaseta/gradcourses/Math785/notes785.html>,

- [2] Andrei Borisvich Shidlovskii:
Transcendental Numbers,
Berlin . New York, Walter de Gruyter, ('89)

- [3] Intel Technology Journal: The computation of Transcendental Functions
<http://developer.intel.com/technology/itj/q41999/articles/art5.htm>,

- [4] Fundamentals of computer science II: IEEE floating-point representation of real numbers.
<http://www.math.grin.edu/~stone/courses/fundamentals/IEEE-reals.html>,

- [5] Definition of Liouville Number
<http://br.crashed.net/~krowne/crc/math/l/l1332.htm>