PMATH 940: HEIGHTS AND ARITHMETIC

HEESUNG YANG

1. September 14

This class will study *heights and arithmetic*. To put it simply, a height function measures the 'complexity' of an algebraic number. In particular, we will explore the application of height functions to Diophantine approximation. Pre-requisites include some algebraic number theory; there shall be no text.

Definition 1. An *algebraic number* is the root of a non-zero irreducible polynomial with integer coefficients. If it is the root of a non-zero monic irreducible with integer coefficients, then it is said to be an *algebraic integer*.

Given an algebraic number α, we would like to measure the "complexity" of α. One such measure is the *height*. In fact, there are several height functions.

Start with an integer a. The measure of complexity is $|a|$. How about rationals? Consider $a/b \in \mathbb{Q}$ where $a, b \in \mathbb{Z}$. A possible suggestion might be $\max(|a|, |b|)$. But this is not well-defined; but this can be circumvented by stipulating that $\gcd(a, b) = 1$.

Definition 2. Suppose $f(x) \in \mathbb{Z}[x]$. Then the *content* of f is the greatest common divisor (GCD) of the integer coefficients of f.

Now let's try to generalize this for a general algebraic number α. For any α, we can associate it with a minimal polynomial f over the integers where α is a root of f, f is of minimal degree with this property, f has content 1, and the leading coefficient is positive. This completely defines f. Say $f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$. Now we defined the *naïve* height of α:

Definition 3. The "*naïve height*" of α, which we denote $H_0(\alpha)$, is defined to be $H_0(\alpha) := \max(|a_d|, \ldots, |a_0|)$.

Notice that if $\alpha = a/b$ with a and b coprime integers and $b > 0$, then the minimal polynomial of a/b over the integers is $f(x) = bx - a$, so $H_0(a/b) = \max(|a|, |b|)$.

Remark 1. Notice that there are only finitely many algebraic numbers of naïve height below any given bound if we restrict the degree below another bound (special case of Northcott's theorem).

Definition 4. Let α be an algebraic number and let f be the minimal polynomial of α over the integers. Suppose also that f factors over $\mathbb{C}$ as

$$f(x) = a_d \prod_{i=1}^{d} (x - \alpha_i).$$

Date: 25 November 2015.

We then define the *Mahler measure* $\mathcal{M}(\alpha)$ by

$$\mathcal{M}(\alpha) = |a_d| \cdot \prod_{i=1}^{d} \max(1, |\alpha_i|).$$

Theorem 1 (Jensen's formula, from Wikipedia). *Suppose that f is an analytic function in a region in the complex plane which contains the closed disk D of radius r about the origin, $a_1, a_2, \cdots, a_n$ are the zeros of f in the interior of D repeated according to multiplicity, and $f(0) \neq 0$. Then*

$$\log|f(0)| = \sum_{k=1}^{n} \log\left(\frac{|a_k|}{r}\right) + \frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\, d\theta.$$

Remark 2. By Jensen's formula, the Mahler measure can be written in a different manner:

$$\mathcal{M}(\alpha) = \exp\left(\int_0^1 \log|f(e^{i\theta})|\, d\theta\right).$$

Definition 5. We define the *absolute Weil height* $H(\alpha)$ of α by

$$H(\alpha) := (\mathcal{M}(\alpha))^{1/d}.$$

The *absolute logarithmic Weil height of* α, denoted $h(\alpha)$, is defined by

$$h(\alpha) := \log H(\alpha) = \frac{1}{d}\log \mathcal{M}(\alpha).$$

The Weil height $H(\alpha)$ is more "natural" and has nicer properties than the naïve height. How so? We will take a small detour to qualify this statement further.

One reason is that there is an alternative definition of the Weil height in terms of valuations on the field $k = \mathbb{Q}(\alpha)$.

Definition 6. An *absolute value* on a field k is a function $|\_| : k \to \mathbb{R}_{\geq 0}$, satisfying

(i) $|x| = 0 \Leftrightarrow x = 0$
(ii) $|xy| = |x| \cdot |y|$ for all $x, y \in k$
(iii) $|x + y| \leq |x| + |y|$ for all $x, y \in k$.

If property (iii) can be strengthened to the strong triangle inequality $|x + y| \leq \max(|x|, |y|)$, then we say that this absolute value is *non-Archimedean*. Otherwise, it is said to be *Archimedean*.

For any absolute value $|\_|$ on a field k, we can introduce a distance function $d(x, y)$ that measure the distance between x and y, by putting $d(x, y) := |x - y|$ for all $x, y \in k$, making k a metric space under d, thereby inducing a topology.

Definition 7. Any absolute values that induce the same topology are said to be *equivalent*.

Definition 8. For the sake of completeness (even though this is totally an uninteresting absolute value): on any field k we have the *trivial absolute value* $|\_|_0$ given by

$$|x|_0 := \begin{cases} 1 & (x \neq 0) \\ 0 & (x = 0). \end{cases}$$

On $\mathbb{Q}$ the ordinary absolute value $|\_|$ is an absolute value. Further, for each prime p we can define an absolute $|\_|_p$ in the following way. For each non-zero integer a, we define $\operatorname{ord}_p a$ to be the exact power of p dividing a. We extend the order function to the rationals by putting

$$\operatorname{ord}_p\left(\frac{a}{b}\right) = \operatorname{ord}_p a - \operatorname{ord}_p b.$$

Further, we define $|\_|_p$ on $\mathbb{Q}$ by

$$|x|_p = p^{-\operatorname{ord}_p x}$$

for $x \neq 0$ and $|0|_p = 0$. Then $|\_|_p$ is an absolute value.

Definition 9. The $|\_|_p$ as defined above is said to be the p-adic absolute value.

Theorem 2 (Ostrowski's theorem)**.** *Every* non-trivial *valuation on $\mathbb{Q}$ is equivalent to the ordinary absolute value or to $|\_|_p$ for some prime p.*

2. September 16

Recall Ostrowski's theorem: every non-trivial absolute value on $\mathbb{Q}$ is equivalent to the ordinary absolute value or to a p-adic absolute value $|\_|_p$ for some prime p. We have the product formula in $\mathbb{Q}$: for each non-zero x in $\mathbb{Q}$, i.e.,

$$|x| \prod_{p \text{ prime}} |x|_p = 1.$$

Notice that if $x = a/b$ with a and b coprime integers with $b > 0$ then

$$H(x) = \max(|b|, |a|) = |b| \max\left(1, \frac{|a|}{|b|}\right) = \max\left(1, \frac{|a|}{|b|}\right) \cdot \prod_p \max\left(1, \left|\frac{a}{b}\right|_p\right) = \prod_\nu \max(1, |x|_\nu),$$

where v runs over the set of normalized inequivalent valuations ν on $\mathbb{Q}$. So

$$H(x) = \prod_\nu \max(1, |x|_\nu)$$

for $x \neq 0$. Now let's turn our attention to algebraic numbers. Let k be a finite extension of $\mathbb{Q}$ so $k = \mathbb{Q}(\alpha)$ for some algebraic number α. Suppose that $[k : \mathbb{Q}] = d$ and there are r real embeddings of k in $\mathbb{C}$ and s pairs of non-real embeddings of k in $\mathbb{C}$. We have $d = r + 2s$. We define $r + s$ valuation ν on k which are related to the ordinary absolute value on $\mathbb{Q}$. Let σ be an embedding of k in $\mathbb{C}$ which maps into $\mathbb{R}$. We then define the valuation ν on k by

$$|\beta|_\nu := |\sigma(\beta)|^{1/d}$$

for any $\beta \in k$. Similarly, if σ is a non-real embedding we defined ν by

$$|\beta|_\nu = |\sigma(\beta)|^{2/d}$$

for any $\beta \in k$. Up to equivalence, there are no other Archimedean valuations that are non-trivial (though the trivial one is non-Archimedean). What about non-Archimedean absolute values ν?

For each prime p we can ask how the principal ideal generated by p in the ring of algebraic integers $\mathcal{O}_k$ of k factors. Say $(p) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_t^{e_t}$ with $\mathfrak{p}_1, \ldots, \mathfrak{p}_t$ prime ideals of $\mathcal{O}_k$ and $e_1, e_2, \ldots, e_t \in \mathbb{Z}_+$. For any ideal A in $\mathcal{O}_k$ let $\mathrm{N}(A)$ denote the norm of A. Then

$$p^d = \mathrm{N}((p)) = \prod_{i=1}^{t} \mathrm{N}(\mathfrak{p}_i)^{e_i} = p^{e_1 f_1 + \cdots + e_t f_t}$$

for appropriate $f_1, \ldots, f_t \in \mathbb{Z}_+$. For any non-zero ideal A in $\mathcal{O}_k$ and any prime ideal $\mathfrak{p}$ we define $\operatorname{ord}_{\mathfrak{p}} A$ to be the exact power of $\mathfrak{p}$ dividing A. By considering the principal ideal generated by x in $\mathcal{O}_k$ we define $\operatorname{ord}_{\mathfrak{p}} x$ for $x \in \mathcal{O}_k$. By taking differences and considering fractional ideals we can extend $\operatorname{ord}_{\mathfrak{p}}$ to all $x \in k^* := k \setminus \{0\}$.

Definition 10. We define *a valuation ν associated with a prime ideal* $\mathfrak{p}$ by

$$|\beta|_\nu = \mathrm{N}(\mathfrak{p})^{-\frac{\operatorname{ord}_{\mathfrak{p}} \beta}{d}}.$$

Again, it can be shown that this defines a non-Archimedean valuation on k. This gives all the non-trivial non-Archimedean valuations up to equivalence. By our normalizations we once again have the product formula

$$\prod_\nu |x|_\nu = \begin{cases} 1 & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

Here, the product is taken over normalized inequivalent valuations ν. Further, we have for $x \in k$ that

$$H(x) = \prod_\nu \max(1, |x|_\nu). \tag{1}$$

By our construction, the height function is properly defined not just on a fixed field k, but it is invariant under finite extensions and so is well-defined on $\overline{\mathbb{Q}}$ (the algebraic closure of $\mathbb{Q}$) also. Thus $H : \overline{\mathbb{Q}} \to \mathbb{R}$. Note that all algebraic closures of $\mathbb{Q}$ are isomorphic. We therefore see from (1) that for any positive integer n and any algebraic number β we have

$$H(\beta^n) = H(\beta)^n.$$

By the product formula, it follows

$$H(\beta) = \prod_\nu \max(1, |\beta|_\nu) = \prod_\nu \max(1, |\beta^{-1}|_\nu) = H(\beta^{-1}).$$

Therefore, for every integer n we have $H(\beta^n) = H(\beta)^{|n|}$. Notice that if ζ_n is an n-th root of unity then $H(\zeta_n) = 1$. In fact, if β is a non-zero algebraic number which is not a root of unity then $H(\beta) > 1$. This was proved by Kronecker in 1857.

3. September 18

Recall that if β is a root of unity or $\beta = 0$ then $H(\beta) = 1$.

Theorem 3. *If β is a non-zero algebraic number with $H(\beta) = 1$ then β is a root of unity.*

Proof. Suppose that β is a non-zero algebraic number with $H(\beta) = 1$. Let

$$f(x) = a_d x^d + \cdots + a_1 x + a_0$$

be the minimal polynomial of β over the integers. Suppose

$$f(x) = a_d \prod_{i=1}^{d} (x - \beta_i),$$

where $\beta = \beta_1, \dots, \beta_d$ are the conjugates of β. Since $H(\beta) = 1$ then we see that $a_d = 1$. Thus β is an algebraic integer. Also, indeed $|\beta_i| \leq 1$ for all $i = 1, 2, \dots, d$. We now consider the algebraic integers $\beta, \beta^2, \dots, \beta^t, \dots$. Note that β^t is a root of the polynomial $f_t(x)$, where

$$f_t(x) := \prod_{i=1}^{d} (x - \beta_i^t).$$

Notice that the conjugates β^t are in the collection $\beta_1^t, \dots, \beta_d^t$, perhaps with repetition. In particular we see that $f_t(x) \in \mathbb{Z}[x]$. The integer coefficients of t_t are elementary symmetric polynomials in $\beta_1^t, \dots, \beta_d^t$. Since $|\beta_i| \leq 1$ we have $|\beta_i^t| \leq 1$ for $i = 1, \dots, d$. Thus the coefficients of f_t are at most 2^d in absolute value. Further each polynomial has at most d roots. Therefore there are positive integers t_1 and t_2 with $t_1 < t_2$ for which $\beta^{t_1} = \beta^{t_2}$. Note $\beta \neq 0$ so $\beta^{t_2 - t_1} = 1$. Therefore β is a root of unity. □

Another proof. On noting that the powers of $\beta, \beta^2, \beta^3, \dots$ of β lie in the unit disc, we can find two powers t_1 and t_2 with $t_2 > t_1$ such that $|\beta^{t_1} - \beta^{t_2}| < 2^{-d}$. Notice that the conjugates $\beta_i^{t_1} - \beta_i^{t_2}$ of $\beta^{t_1} - \beta^{t_2}$ satisfy $|\beta_i^{t_1} - \beta_i^{t_2}| \leq 2$. But $\beta = \beta_1$ and

$$\prod_{i=1}^{d} (\beta_i^{t_1} - \beta_i^{t_2}) \tag{2}$$

is an integer. Note also that the absolute value of (2) is less than 1, so (2) is equal to 0. Hence $\beta^{t_1} = \beta^{t_2}$ and so either $\beta = 0$ or β is a root of unity. □

If β is an algebraic number with $\beta \neq 0$ and β not a roof of unity then $H(\beta) > 1$. So the natural question: is there a real number $\varepsilon > 0$ such that if β is an algebraic number and $H(\beta) < 1 + \varepsilon$ then $H(\beta) = 1$? The answer is no; for this, we need a sequence of numbers whose height approaches 1. Take $\beta = 2^{1/n}$. Then each of the conjugates β_i of β satisfy $|\beta_i| = 2^{1/n}$. There are n such conjugates, i.e., the degree of β is n. Thus

$$H(\beta) = \left(\prod_{i=1}^{n} \max(1, 2^{1/n}) \right)^{1/n} = 2^{1/n},$$

and $2^{1/n} \to 1$ as $n \to \infty$. However, if we ask the same question for the Mahler measure $\mathcal{M}(\beta)$ we don't know the answer. This is known as Lehmer's question. He posed it in 1933. Recall that $\mathcal{M}(\beta) = H(\beta)^d$ where d is the degree of β. Also we have $\mathcal{M}(\beta) = cM(f)$ where f is the minimal polynomial of β. So Lehmer's question can be stated as follows: Does there exists $\varepsilon_0 > 0$ such that $\mathcal{M}(f) < 1 + \varepsilon_0$ implies $\mathcal{M}(f) = 1$, or if $\mathcal{M}(\beta) < 1 + \varepsilon_0$ then $\mathcal{M}(\beta) = 1$?

Lehmer gave an example of an algebraic number with Mahler measure larger than 1 but small. His example: let β be the largest real root of $f(x)$ where $f(x) = x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1$. f is irreducible over $\mathbb{Q}$ and $\mathcal{M}(f) = \mathcal{M}(\beta) = \beta = 1.17628081\dots$.

4. September 21

Recall Lehmer's question from 1933: is there a positive number ε_0 such that if $f \in \mathbb{Z}[x]$ with $\mathcal{M}(f) < 1 + \varepsilon_0$ then $\mathcal{M}(f) = 1$? Lehmer gave a possible candidate for $\varepsilon_0 = .17628081\dots$.

Definition 11. A *reciprocal polynomial* is a polynomial f of degree d for which $f(x)$ is identically equal to $x^d f(x^{-1})$. Equivalently, for each root α of f, α^{-1} is also a root of f.

In 1971, C. Smyth proved that Lehmer's question has a positive answer if we restrict f to be a polynomial which is not reciprocal. In particular, Smyth proved that if f is a non-reciprocal polynomial which does not have 0 or 1 as a root, then

$$\mathcal{M}(f) \geq \beta_0,$$

where β_0 is the real root of $x^3 - x - 1$. In this case, we have $\beta_0 = 1.3247\dots$.

Theorem 4 (Smyth)**.** *Let $f \in \mathbb{Z}[x]$ be a non-reciprocal polynomial which does not have* 0 *or* 1 *as a root. Then $\mathcal{M}(f) \geq \frac{\sqrt{5}}{2}$.*

Proof. We may suppose that $\mathcal{M}(f) < 2$ since the result is true otherwise, and thus f is a monic polynomial. Further, since $\mathcal{M}(f_1 f_2) = \mathcal{M}(f_1)\mathcal{M}(f_2)$, since $\mathcal{M}(f) \geq 1$ for $f \in \mathbb{Z}[x]$ and since the product of reciprocal polynomials is reciprocal, we may suppose that f is irreducible. Suppose that f is of degree n and that $\alpha_1, \dots, \alpha_n$ are the roots of f. Put $r(x) = x^n f(x^{-1})$. Since f is non-reciprocal and does not have 0 or 1 as a root, $f(x)/r(x)$ is not a constant. We can expand $f(x)/r(x)$ as a power series, say

$$\frac{f(z)}{r(z)} = a_0 + a_k z^k + a_l z^l + \cdots \tag{3}$$

where the coefficients a_0, a_k, a_l are non-zero integers. The coefficients are integers since the constant coefficient of $r(z)$ is equal to 1; recall this follows since $\mathcal{M}(f) < 2$ hence is monic. Further $f(0) = a_0 r(0)$ and so $|a_0| = 1$. We now remark that f has no roots on the unit circle. For suppose α is a root of f on the unit circle. Then $\alpha\bar{\alpha} = 1$. But then for any conjugate α_i of α also satisfies $\alpha_i \overline{\alpha_i} = 1$.

But for every root α of f, α^{-1} is also a root hence f is reciprocal or has 1 as a root. We may not put

$$g(z) = \prod_{|\alpha_j|<1} \left(\frac{z - \alpha_j}{1 - \overline{\alpha_j} z} \right) = c + c_1 z + c_2 z^2 + \cdots \tag{4}$$

and put

$$h(z) = \prod_{|\alpha_j|>1} \left(\frac{1 - \overline{\alpha_j} z}{z - \alpha_j} \right) = d + d_1 z + d_2 z^2 + \cdots . \tag{5}$$

Observe that $\frac{f(z)}{r(z)} = \frac{g(z)}{h(z)}$. Upon comparing (3), (4), and (5) we find that $a_k d + a_0 d_k = c_k$ so $a_k d + d_k = c_k$. Since a_k is a non-zero integer with $|a_k| \geq 1$, it follows that

$$\max(|d_k|, |c_k|) \geq \frac{|d|}{2}. \tag{6}$$

Both g and h have no poles on or inside the unit circle and so they are holomorphic in an open set containing the unit disc. By Parseval's inequality, it follows

$$\frac{1}{2\pi}\int_0^{2\pi} |g(e^{i\theta})|^2 \, d\theta = |c|^2 + |c_1|^2 + |c_2|^2 + \cdots .$$

Since g has absolute value 1 on the unit circle, we have $1 = |c|^2 + |c_1|^2 + \cdots$. Hence $|c_k|^2 \leq 1 - |c|^2$. Similarly we find that $|d_k|^2 \leq 1 - |d|^2$. But $c = d = \mathcal{M}(f)^{-1}$. Therefore from (6) we see that

$$\begin{aligned} &\frac{|d|^2}{4} \leq 1 - |d|^2 \\ \Rightarrow &\frac{5}{4} \leq |d|^{-2} = \mathcal{M}(f)^{-2} \\ \Rightarrow &\mathcal{M}(f) \geq \frac{\sqrt{5}}{2}, \end{aligned}$$

as desired. □

5. September 23

Recall Smyth's theorem which states that if $f \in \mathbb{Z}[x]$ is non-reciprocal and doesn't have 0 or 1 as a root then $\mathcal{M}(f) \geq \beta_0$ where β_0 is the real root of $x^3 - x - 1$. β_0 is an example of a Pisot or Pisot-Vijayaraghavan number (P.V. for short).

Definition 12. A real algebraic integer β is said to be a *Pisot number* if $\beta > 1$ and all other conjugates of β have absolute value **less** than 1.

The set of such numbers is usually denoted by S. It is a closed set (in the usual topology embedded in $\mathbb{R}$; note that this is not obvious, and it will be fairly difficult to prove.). It contains all the integers larger than 1. The Pisot numbers were first studied by Thue in 1912 and by Hardy in 1919.

Observe that if β is a Pisot number then $\mathcal{M}(\beta) = \beta$. Further, β is a root of a non-reciprocal polynomial if the degree of β exceeds 2. It then follows from Smyth's result that β_0 is the smallest Pisot number. This fact was first proved by Siegel in 1944. The smallest non-isolated limit point of S was shown by Dufresnay and Pisot in 1955 to be $\frac{1+\sqrt{5}}{2}$.

For any real number x let $\|x\|$ denote the distance from x to the nearest integer. Let λ be a real number with $\lambda > 1$. We can consider the sequence $(\|\lambda^n\|)_{n=1}^{\infty}$. In general we would expect the sequence to be uniformly distributed in $(0, 1)$. However if λ is a Pisot number then $\|\lambda^n\| \to 0$ as $n \to \infty$. To see this let $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_d$ be the conjugates of λ. Since $\lambda \in S$ we see that $|\lambda_i| < 1$ for $i = 2, \ldots, d$. But $\operatorname{tr}(\lambda^n)$ for $n \in \mathbb{Z}$ is an integer so $\|\lambda^n + \lambda_2^n + \cdots + \lambda_d^n\| = 0$. It then follows from Smyth's result that β_0 is the smallest Pisot number. This fact was first proved by Siegel in 1944. Thus

$$\|\lambda^n\| \leq |\lambda_2|^n + \cdots + |\lambda_d|^n$$

and

$$|\lambda_2|^n + \cdots + |\lambda_d|^n \to 0$$

as $n \to \infty$.

Hence a natural question (still open!): are the Pisot numbers the only real numbers greater than 1 with this property? However it is known that if $\lambda \in \mathbb{R}$ and $\lambda > 1$ and λ is algebraic then $\|\lambda^n\| \to 0$ implies $\lambda \in S$. Further, Pisot proved in 1939 that if $\lambda > 1$ is a real number and

$$\sum_{n=1}^{\infty} \|\lambda^n\| < \infty$$

then λ is a Pisot number. The Pisot numbers arise in several settings.

Definition 13. A subset T of unit circle is said to be a *set of uniqueness* if any trigonometric expansion

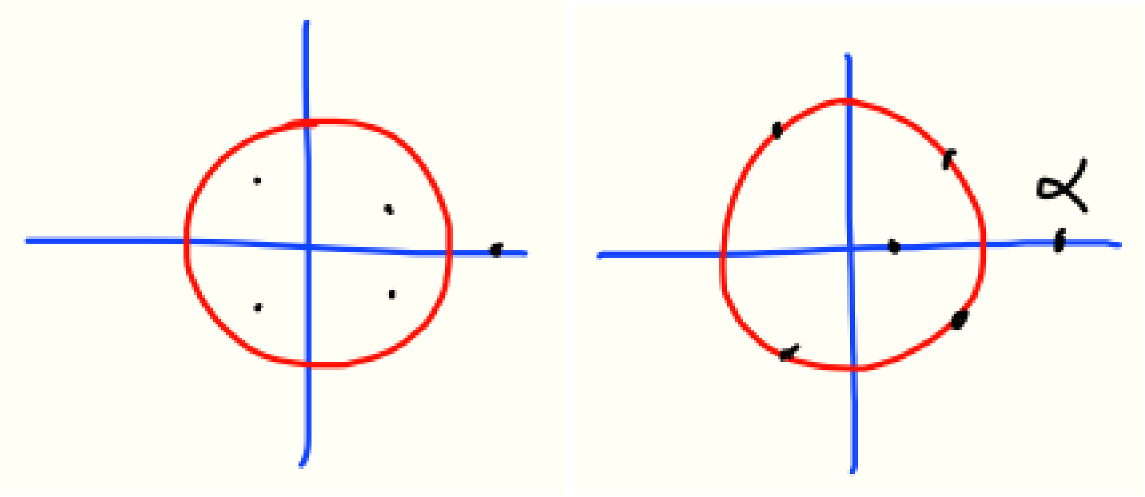
$$\sum_{n=-\infty}^{\infty} c(n)e^{itn}$$

which converges to zero for all t on the unit circle with $t \notin T$ converges to zero everywhere on the unit circle.

Salem and Zygmund proved that if T is a Cantor set of constant ratio of dissection θ on the unit circle then P is a set of uniqueness if and only if θ^{-1} is a Pisot number.

Definition 14. A *Salem number* α is a real algebraic integer with $\alpha > 1$ which has all its other conjugates on or inside the unit circle.

In fact, this implies that α has one real conjugate which is α^{-1} and the other conjugates come in complex conjugate pairs and lie on the unit circle *and at least one on the unit circle.*



The left diagram describes the Pisot number and the right diagram describes the Salem number.

6. September 25

The set of Salem numbers is not so well understood. It is known that every Pisot number is a limit, both from above and below, of a sequence of Salem numbers.

It is not known if this is a smallest Salem number. In 1977 Boyd gave a way to produce all Salem numbers. He found four Salem numbers smaller than 1.22 and conjectured that they are the four smallest. Two of them are of degree 14 and of degree 18. Lehmer's example is of degree 10.

Back to Lehmer's questino: in 1971 Blanksby and Montogomery used Fourier analysis to prove that if θ is a non-zero algebraic number of degree d which is not a root of unity then

$\mathcal{M}(\theta) \geq 1 + \frac{1}{52d \log 6d}$. In 1978 Stewart found an argument from transcendency theory that gave, under the same assumptions, for $d > 1$, $\mathcal{M}(\theta) \geq 1 + \frac{1}{10^4 d \log d}$. In 1979 Dobrovolski extended this approach to prove that if θ is a non-zero algebraic number with degree d and θ is not a root of unity then for each $\varepsilon > 0$,

$$\mathcal{M}(\theta) \geq 1 + (1 - \varepsilon) \left(\frac{\log \log d}{\log d} \right)^3$$

for d sufficiently large in terms of ε. This is the best result as a function of d known to date.

For the transcendence apporach, we require a result on solutions of systems of linear equations known as *Siegel's lemma.* We will prove the following version of Siegel's lemma.

Lemma 1 (Siegel's lemma)**.** *Let $b_{ij} (1 \leq i \leq N, 1 \leq j \leq M)$ be algebraic integers in a field K and suppose that for each j with $1 \leq j \leq M$ not all the b_{ij}'s zero. Suppose that $[K : \mathbb{Q}] = d$ and let $\sigma_1, \ldots, \sigma_d$ be the embeddings of K into $\mathbb{C}$. If $N \geq 2dM$ then the system of equations*

$$\sum_{i=1}^{N} b_{ij} x_i = 0$$

for $j = 1, \ldots, M$, has a solution in rational integers $x_1, \ldots, x_N$ not all of which are zero, where absolute values are at most

$$\sqrt{2} N \left(\max_{i \leq j \leq M} \prod_{k=1}^{d} \left(\max_{1 \leq i \leq N} |\sigma_k(b_{ij})| \right) \right)^{1/d}. \tag{7}$$

Proof. Let $\sigma_1, \ldots, \sigma_r$ be the real embeddings of K in $\mathbb{C}$ and let $\sigma_{r+1}, \ldots, \sigma_d$ be the non-real embeddings of K in $\mathbb{C}$ with

$$\sigma_{r_1+r_2+i} = \overline{\sigma_{r_1+i}}$$

for $i = 1, \ldots, r_2$, where $d = r_1 + 2r_2$.

Put

$$\tau_i = \begin{cases} \sigma_i & \text{for } 1 \leq i \leq r \\ \mathrm{R}e(\sigma_i) & \text{for } r_1 < i \leq r_1 + r_2 \\ \mathrm{I}m(\sigma_i) & \text{for } r_1 + r_2 < i \leq d. \end{cases}$$

Here Re denotes the real part of a complex number and Im the imaginary part. Define Y to be the integer part of expression (7). For any pair of integers (k, j) with $1 \leq k \leq d$ and $1 \leq j \leq M$, the $(Y+1)^N$ different N-tuples $(y_1, \ldots, y_N)$ with $0 \leq y_i$‘Y for $1 \leq i \leq N$ give rise to the numbers $\left| \tau_k \left(\sum_{i=1}^{N} b_{ij} y_i \right) \right|$ which are at most $NY \max_{1 \leq i \leq N} |\tau_k(b_{ij})|$. Put $L := Y(Y+1)$ and observe that L is a non-zero integer, since the b_{ij}'s are algebraic integers which are not all zero. Since $N \geq 2dM$ and $L \leq (Y+1)^2$, we have

$$L^{Md} < (Y+1)^N.$$

Therefore by the pigeonhole principle, two of the N-tuples $(y_1, \ldots, y_N)$ and $(y'_1, \ldots, y'_N)$ satisfy

$$\left| \tau_k \left(\sum_{i=1}^{N} b_{ij} y_i \right) - \tau_k \left(\sum_{i=1}^{N} b_{ij} y'_i \right) \right| \leq \max_{1 \leq i \leq N} |\tau_k(b_{ij})| \frac{NY}{L} \tag{8}$$

for $k = 1, 2, \dots, d$ and $j = 1, \dots, M$. Put $x_i = y_i - y_i'$ for $i = 1, \dots, N$. Then $\max\limits_{1 \le i \le N} |x_i| \le Y$ and not all the x_i's are zero. So it remains to show that

$$\sum_{i=1}^{N} b_{ij} x_i = 0$$

for $j = 1, \dots, M$. From (8), we deduce that

$$\left| \sigma_k \left(\sum_{i=1}^{N} b_{ij} x_i \right) \right| \le \max_{1 \le i \le N} |\sigma_k(b_{ij})| \frac{NY}{L}$$

for $k = 1, \dots, r$ and that

$$\left| \sigma_k \left(\sum_{i=1}^{N} b_{ij} x_i \right) \sigma_{k+r_2} \left(\sum_{i=1}^{N} b_{ij} x_i \right) \right| \le \left\{ \max_{1 \le i \le N} (\mathrm{Re}\, \sigma_k(b_{ij}))^2 + \max_{1 \le i \le N} (\mathrm{I}m\, \sigma_k(b_{ij})) \right\}^2 \left(\frac{NY}{L} \right)^2$$
$$\le 2 \max_{1 \le i \le N} |\sigma_k(b_{ij}) \sigma_{k+r_2}(b_{ij})| \left(\frac{NY}{L} \right)^2$$

for $k = r_1 + 1, \dots, r_1 + r_2$. Therefore

$$\left| \prod_{k=1}^{d} \sigma_k \left(\sum_{i=1}^{N} b_{ij} x_i \right) \right| < \left(\frac{Y(Y+1)}{L} \right)^d = 1$$

for $j = 1, \dots M$. Notice that the above inequality is just $\left| \mathrm{N}_{K/\mathbb{Q}} \left(\sum\limits_{i=1}^{N} b_{ij} x_i \right) \right|$ and since it is less than 1 it is necessarily 0. Thus $\sum\limits_{i=1}^{N} b_{ij} x_i = 0$ for $j = 1, \dots, M$ as required. □

7. September 28 & 30

Theorem 5. *If α is a non-zero algebraic integer of degree $d > 1$ with*

$$\mathcal{M}(\alpha) < 1 + (10^4 d \log d)^{-1}$$

then α is a root of unity.

Proof. It is easy to check that this holds for $d = 2, 3$. So we may assume that $d \ge 4$. Let $\alpha = \alpha_1, \dots, \alpha_d$ be the conjugates of α. We may assume without loss of generality that

$$|\alpha| \ge |\alpha_i| \quad (i = 1, \dots, d). \tag{9}$$

Put

$$K := 2U, \text{ where } U = \lfloor 70 d \log d \rfloor .$$

We now choose K positive integers $r_1 < r_2 < \cdots < r_k$ from the first $13K$ positive integers so that

$$\max_{1 \le s \le t \le K} \{ |\mathrm{I}m(\log \alpha^{r_s}) - \mathrm{I}m(\log \alpha^{r_t})| \} \le \frac{2\pi}{13}. \tag{10}$$

Here $\mathrm{I}m(z)$ denotes the imaginary part of z for $z \in \mathbb{C}$, and $\log z$ denotes the principal branch of the logarithm function where $-\pi < \mathrm{I}m(\log z) \le \pi$. Such a choice is possible by the pigeonhole principle. Put $\theta_1 = \min\limits_{1 \le j \le K} \mathrm{I}m(\log \alpha^{r_j})$ and put $\theta = \theta_1 + \frac{\pi}{13}$.

We now construct a function $f(z)$ where

$$f(z) = \exp(-i\theta z) \sum_{k=1}^{K} \sum_{j=1}^{d} a_{k,j} \alpha^j \exp((\log \alpha^{r_k})z),$$

and where the $a_{k,j}$'s are rational integers (not all zero) which are chosen so that $f(u) = 0$ for $u = 1, 2, \ldots, U$. Notice that this in equivalent to solving

$$\sum_{k=1}^{K} \sum_{j=1}^{d} a_{k,j} \alpha^{j+r_k u} = 0,$$

for $u = 1, \ldots, U$. Since Kd, the number of unknowns, is $2d \cdot U$, the number of equations, we may apply Siegel's lemma to get a non-trivial solution in rational integers $a_{k,j}$ with

$$\max_{k,j} |a_{k,j}| \leq \sqrt{2} K d M^{13KU+d}. \tag{11}$$

where

$$M = \left(\prod_{\sigma \in S} \max(1, |\sigma(\alpha)|) \right)^{1/d} = (\mathcal{M}(\alpha))^{1/d};$$

S denotes the set of embeddings of $\mathbb{Q}(\alpha)$ in $\mathbb{C}$.

We now define f in terms of the $a_{k,j}$'s. We shall now show that f has more zeros. In particular, we prove by induction that $f(u) = 0$ for all positive integers u. Accordingly suppose that $f(u) = 0$ for $u = 1, 2, \ldots, J$ with $J \geq U$. And we will prove that $f(J+1) = 0$. Since f is an entire function so is F, where

$$F(z) := \frac{f(z)}{(z-1)(z-2)\cdots(z-J)}.$$

By the maximum modulus principle,

$$|F(J+1)| \leq \max_{z \in \Gamma} |F(z)|,$$

where

$$\Gamma = \{z : z \in \mathbb{C}, |z| = 2J+1\}.$$

Thus

$$|f(J+1)| \leq J! \max_{z \in \Gamma} \left(\prod_{j=1}^{J} (z-u) \right)^{-1} \max_{z \in \Gamma} |f(z)|$$

$$|f(J+1)| \leq \frac{J! J!}{(2J)!} \max_{z \in \Gamma} |f(z)|$$

$$|f(J+1)| \leq \binom{2J}{J}^{-1} \max_{z \in \Gamma} |f(z)|. \tag{12}$$

We now estimate $\max_{z \in \Gamma} |f(z)|$. On recalling (9) and (11) we see that

$$\max_{z \in \Gamma} |f(z)| \leq \sqrt{2} (Kd)^2 M^{13KU+d} |\alpha|^d \exp(\Delta(2J+1)) \tag{13}$$

where $\Delta = \max_{1\le k\le K} |\log(\alpha^{r_k}) - i\theta|$. By (10), we see that

$$\Delta \le \left|13K\log|\alpha| + i\frac{\pi}{13}\right|.$$

Since $|\alpha| \ge |\alpha_i|$ for $i = 1,\ldots,d$ we may use the fact that $1 \le |\alpha| \le \mathcal{M}(\alpha)$ our assumption that $1 \le \mathcal{M}(\alpha) \le 1 + (10^4 d\log d)^{-1}$ and the inequality that $\log(1+x) \le x$ for $x \ge 0$ to conclude that $0 \le \log|\alpha| \le (10^4 d\log d)^{-1}$. Now since $K = 2U$ and $U = \lfloor 70d\log d\rfloor$, we see that $0 \le 13K\log|\alpha| \le \frac{\pi}{13}$; and so $\Delta \le \frac{\sqrt{2}\pi}{13}$, whence $\Delta(2J+1) \le J(\log 2)$.

Recall that

$$f(z) = \exp(-i\theta z)\sum_{k=1}^{K}\sum_{j=1}^{d} a_{k,j}\alpha^j \exp((\log\alpha^{r_k})z).$$

We chose the $a_{k,j}$'s to be integers which are not too large so that $f(u) = 0$ for $u = 1, 2, \ldots, U$.

We are now proving by induction that $f(u) = 0$ for all positive integers u. We then have from (12) and (13) that

$$|f(J+1)| \le \binom{2J}{J}^{-1} 2^J\sqrt{2}(Kd)^2 M^{13KU+d}|\alpha|^d.$$

Since $\binom{2J}{J} \ge \frac{4^J}{2J}$ we have

$$|f(J+1)| \le 2J\cdot 2^{-J}\sqrt{2}(Kd)^2 M^{13KU+d}|\alpha|^d,$$

so

$$|f(J+1)| \le J2^{-J}K^4M^{26KU}. \tag{14}$$

Our next step is to show that $|f(J+1)|$ is so small that it must be zero.

We now estimate $|f(J+1)|$ from below. Put $\beta = f(J+1)\exp(i\theta(J+1))$ and notice that β is an algebraic integer in $\mathbb{Q}(\alpha)$. Therefore either $\beta = 0$ in which case $f(J+1) = 0$ or $|\operatorname{N}_{\mathbb{Q}(\alpha)/\mathbb{Q}}(\beta)|$ is a positive integer hence

$$|f(J+1)| = |\beta| \ge \left(\prod_{\sigma\in S'}|\sigma(B)|\right)^{-1} \tag{15}$$

where S' is the set of embeddings S' of $\mathbb{Q}(\alpha)$ in $\mathbb{C}$ minus the identity embedding. Notice that for all $\sigma \in S$ we have

$$|\sigma(B)| \le \sqrt{2}(Kd)^2M^{13KU+d}\max\left(1, |\sigma(\alpha)|^{13K(J+1)+d}\right). \tag{16}$$

Since $|\alpha| \ge |\alpha_i|$ for $i = 1, 2, \ldots, d$, we have

$$\prod_{\sigma\in S'}\max(1,|\sigma(\alpha)|) \le \left(\prod_{\sigma\in S}\max(1,|\sigma(\alpha)|)\right)^{\frac{d-1}{d}} = M^{d-1}.$$

So now by (15) and (16), we find that

$$|f(J+1)| \ge (K^4M^{26K(J+1)})^{-d+1}.$$

We compare this estimate with (14) to get

$$2^J \le JK^4M^{26KU}(K^4M^{26K(J+1)})^{d-1}.$$

so

$$2^J \leq JK^{4d}M^{26K(J+1)d}.$$

Taking logarithms and estimating $\frac{J+1}{J}$ from above by $\frac{27}{26}$ we find

$$\log 2 \leq \frac{\log J}{J} + \frac{4d \log K}{J} + 27Kd \log M.$$

Thus, upon recalling that $cM(\alpha) = M^d, K = 2U$ and $J \geq a$, we have

$$\log 2 \leq \frac{\log U}{U} + \frac{4d \log 2U}{U} + 54U \log(\mathcal{M}(\alpha)). \tag{17}$$

Since $U = \lfloor 70d \log d \rfloor$ and $d \geq 4$, we find that

$$\frac{\log U}{U} + \frac{4d \log 2U}{U} < .31.$$

Thus by (17),

$$(\log 2 - 0.31) \leq 54U \log(\mathcal{M}(\alpha)).$$

But $\log(1+x) \leq x$ for $x \geq 0$ and so

$$\log 2 - .31 \leq \frac{54U}{10^4 d \log d}.$$

Hence

$$\frac{(\log 2 - 0.31)10^4 d \log d}{54} \leq U = \lfloor 70d \log d \rfloor.$$

Contradiction! This is false and so $f(J+1) = 0$. Therefore by induction $f(u) = 0$ for $u = 1, 2, 3, \ldots$.

Let us put $A_k := \sum_{j=1}^{d} a_{k,j}\alpha^j$. Then

$$f(u)\exp(i\theta u) = \sum_{k=1}^{K} A_k(\alpha^U)^{r_k}) = 0$$

for $u = 1, 2, \ldots$. Notice that since α has degree d, A_k is zero if and only if $a_{k,j} = 0$ for $j = 1, \ldots, d$. Since not all of the $a_{k,j}$'s are zero we see that not all of the A_k's are zero. Thus $g(x) = \sum_{k=1}^{K} A_k x^{r_k}$ is a non-zero polynomial. But α^U is a root of $g(x)$ for $u = 1, 2, \ldots$. Since $\alpha \neq 0$ we see that $\alpha^{U_1} = \alpha^{U_2}$ for some distinct positive integers and as α is a root of unity. □

Remark 3. We briefly get back to Pisot for a bit. He proved in 1938 that if λ is a real number with $\lambda > 1$ and $\sum_{n=1}^{\infty} \|\lambda^n\|^2 < \infty$, then λ is a Pisot number.

8. October 2

Let α be a real number and suppose that α is irrational.

Question 1 (Basic question)**.** How well can we approximate α by rationals?

Answer 1. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ we can approximate α to within ε for any $\varepsilon > 0$. A better question would be to ask how well we can approximate α in terms of the size of the denominator of the rationals?

Theorem 6 (Dirichlet)**.** *If $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, then there exist infinitely many p/q with $p, q \in \mathbb{Z}, q > 0, (p,q) = 1$ for which $\left|\alpha - \frac{p}{q}\right| > \frac{1}{q^2}$.*

So one may naturally ask if this result is sharp. In fact, it turns out that Dirichlet's result is indeed sharp: there exist $C > 0$ such that for uncountably many $\alpha \in \mathbb{R}$, we have $\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^2}$ for all rationals $\frac{p}{q}, q > 0, (p,q) = 1$.

One more natural question:

Question 2. What happens if we restrict α to be algebraic?

The first interesting response to this question was given by Liouville in 1844.

Theorem 7 (Liouville)**.** *Let α be algebraic of degree $d \geq 2$. Then there exists $C(\alpha) > 0$ such that for all $p/q \in \mathbb{Q}$ with $(p,q) = 1, q > 0$, we have*

$$\left|\alpha - \frac{p}{q}\right| > \frac{C(\alpha)}{q^d}.$$

This result gives us a recipe for constructing transcendental numbers since we need only find an $\alpha \in \mathbb{R}$ with a sequence $(p_i/q_i)_{i=1}^{\infty} \in \mathbb{Q}$ with

$$\left|\alpha - \frac{p_j}{q_j}\right| < \frac{1}{q_j^j}$$

for $j = 1, 2, \ldots$. This is how Liouville constructed the first real number known to be transcendental. The partial sums

$$\frac{p_j}{q_j} = \sum_{k=1}^{j} \frac{1}{10^{k!}}$$

give the required sequence.

Can we improve on Liouville's result when $d \geq 3$? Yes, but even very small improvements seem very difficult to achieve. Thus in 1909 was the first to make an improvement, followed by Siegel in 1921, Dyson in 1947, and finally by Roth in 1955.

Theorem 8 (Roth)**.** *Let α be algebraic of degree $d \geq 2$. Let $\varepsilon > 0$. Then there exists a constant $C(\alpha, \varepsilon) > 0$ (i.e., a constant depending on α and ε) such that*

$$\left|\alpha - \frac{p}{q}\right| > \frac{C(\alpha, \varepsilon)}{q^{2+\varepsilon}}.$$

However, there is a *big flaw* in this extraordinary theorem. The proof does not give a means to compute $C(\alpha, \varepsilon) > 0$ explicitly given α and ε. Thus the result is said to be *ineffective* and it is a major open problem to make it effective.

In general, how do we find the "good" rational approximations to an $\alpha \in \mathbb{R}$?

Definition 15. We say an approximation $\frac{p}{q}$ is *good* if $\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$.

It turns out that there is a very efficient algorithm to find these approximations known as the continued fraction algorithm. There is also a method known as the *hypergeometric method* which gives effective improvement of Liouville's result for certain algebraic numbers such as $\sqrt[3]{2}$. The idea is to consider sequence $\frac{P_n(x)}{Q_n(x)}$ of polynomials of degree at most n which approximate $(1-x)^{1/3}$ and then specialize.

9. October 5

We consider the function of the N-th variables $a_0, \dots, a_N$:

Definition 16. We define the *partial fraction*

$$[a_0, \dots, a_N] := a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_N}}}.$$

We call $a_0, \dots, a_N$ *partial coefficients.*

From the definition we have

$$[a_0] = a_0, [a_0, a_1] = a_1 + \frac{1}{a_1} = \frac{a_0a_1 + 1}{a_1}, [a_0, a_1, a_2] = a_1 + \frac{1}{a_1 + \frac{1}{a_2}} = \frac{a_2a_1a_0 + a_2 + a_0}{a_2a_1 + 1}.$$

We also have

$$[a_0, \dots, a_n] = \left[a_0, a_1, \dots, a_{n-1} + \frac{1}{a_n}\right]$$
$$[a_0, \dots, a_n] = [a_0, [a_1, \dots, a_n]].$$

More generally, $[a_0, \dots, a_n] = [a_0, \dots, a_{m-1}, [a_m, \dots, a_n]]$.

Definition 17. We call $[a_0, \dots, a_n] (0 \le n \le N)$ the n-th convergent to $[a_0, \dots, a_N]$.

Theorem 9. *If p_n and q_n are defined by*

$$\begin{cases} p_0 = a_0, p_1 = a_1a_0 + 1, p_n = a_np_{n-1} + p_{n-2} & (n \ge 2) \\ q_0 = 1, q_1 = a_1, \dots, q_n = a_nq_{n-1} + q_{n-2} & (n \ge 2). \end{cases}$$

Then $[a_0, \dots, a_n] = \frac{p_n}{q_n}$.

Proof. We prove with induction. Clearly, we have $[a_0] = a_0 = \frac{a_0}{1} = \frac{p_0}{q_0}$ and $[a_0, a_1] = \frac{a_0a_1+1}{a_1} = \frac{p_1}{q_1}$. Suppose it is true for $n \le m < N$. Then

$$[a_0, \dots, a_m] = \frac{p_m}{q_m} = \frac{a_mp_{m-1} + p_{m-2}}{a_mq_{m-1} + q_{m-2}}.$$

Also, $p_{m-1}, p_{m-2}, q_{m-1}, q_{m-2}$ depend on $a_0, a_1, \dots, a_{m-1}$. We have

$$\begin{aligned}[a_0, a_1, \dots, a_{m+1}] &= \left[a_0, a_1, \dots, a_m + \frac{1}{a_{m+1}}\right] \\ &= \frac{\left(a_m + \frac{1}{a_{m+1}}\right) p_{m-1} + p_{m-2}}{\left(a_m + \frac{1}{a_{m+1}}\right) q_{m-1} + q_{m-2}} \\ &= \frac{a_{m+1}(a_m p_{m-1} + p_{m-2}) + p_{m-1}}{a_{m+1}(a_m q_{m-1} + q_{m-2} + q_{m-1}} \\ &= \frac{a_{m+1} p_m + p_{m-1}}{a_{m+1} q_m + q_{m-1}} = \frac{p_{m+1}}{q_{m+1}}.\end{aligned}$$

□

Theorem 10. $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$. *Equivalently,*

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

Proof. $p_n q_{n-1} - p_{n-1} q_n = a_n p_{n-1} + p_{n-2} - p_{n-1}(a_n q_{n-1} + q_{n-2}) = -(p_{n-1} q_{n-2} - p_{n-2} q_{n-1}) = (-1)^{n-1}$. □

Theorem 11. $p_n q_{n-2} + p_{n-2} q_n = (-1)^n a_n$, *or*

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

Proof. Exercise. □

From now on, we shall assume that $a_0 \in \mathbb{Z}$ and $a_1, a_2, \dots, \in \mathbb{N}$. Also, let $x_n = \frac{p_n}{q_n}$ and $x = x_N$ be the N-th convergent.

Theorem 12. $x_0 < x_2 < x_4 < \cdots$ *and* $x_1 > x_3 > x_5 > \cdots$.

Proof. This follows from Theorem 11. □

Theorem 13. *every odd convergent i greater than any even convergent. That is,* $x_{2m+1} > x_{2\mu}$ *where* $2\mu, 2m+1 \leq N$.

Proof. From Theorem 10, we have $x_{2m+1} > x_{2m}$. If $\mu \leq m$, then $x_{2m} > x_{2\mu}$ so $x_{2m+1} > x_{2\mu}$. If $\mu > m$, then $x_{2\mu} < x_{2\mu+1}$; since $x_{2m+1} > x_{2\mu+1}$ it follows $x_{2m+1} > x_{2\mu}$. □

Theorem 14. $x = x_N$ *is greater than any even convergent and less than any odd convergent.*

Let α be a real number. We construct a continued fraction associated with α, using following steps:

Step 1: Define $a_0 := \lfloor \alpha \rfloor$. If $\alpha = a_0$ then $\alpha = [a_0]$. Otherwise, then $\alpha = a_0 + \frac{1}{\alpha_1}$ for appropriate α_1.

Step 2: Let $a_1 = \lfloor \alpha_1 \rfloor$. If $\alpha_1 = a_1$ then $\alpha = a_0 + \frac{1}{a_1} = [a_0, a_1]$.

We repeat this procedure. If this stops after a finite number of steps then $\alpha = [a_0, \dots, a_N]$. Otherwise, then $\alpha = [a_0, a_1, \dots]$, an infinite continued fraction.

Remark 4. α has a finite continued fraction if and only if α is a rational number.

Proposition 1. *The sequence* $(|q_1\alpha - p_1|, |q_2\alpha - p_2|, \dots)$ *is a strictly decreasing sequence.*

Proof. Let $\alpha = [a_0, a_1, \dots, a_n, \alpha_{n+1}]$. Then $\alpha = \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}}$. Also,

$$\begin{aligned} |q_n\alpha - p_n| &= \left| q_n \left(\frac{p_n\alpha_{n+1} + p_{n-1}}{q_n\alpha_{n+1} + q_{n-1}} \right) - p_n \right| \\ &= \frac{|q_np_{n-1} - p_nq_{n-1}|}{|q_n\alpha_{n+1} + q_{n-1}|} = \frac{1}{q_n\alpha_{n+1} + q_{n-1}} \\ q_n\alpha_{n+1} + q_{n-1} &\geq q_n + q_{n-1} \geq a_nq_{n-1} + q_{n-2} + q_{n-1} \\ &\geq (a_n + 1)q_{n-1} + q_{n-2} > \alpha_nq_{n-1} + q_{n-2}. \end{aligned}$$

So

$$|q_n\alpha - p_n| = \frac{1}{q_n\alpha_{m-1} + q_{n-1}} < \frac{1}{q_{n-1}\alpha_n + q_{n-2}} = |q_{n-1}\alpha - p_{n-1}|.$$

Thus the sequence is strictly decreasing. □

10. October 7

Proposition 2. $\frac{1}{(a_{n+1}+2)q_n^2} < \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{a_{n+1}q_n^2}$.

Proof. From the proof of Prop 1, we have

$$\left|\alpha - \frac{p_n}{q_n}\right| = \frac{1}{q_n(\alpha_{n+1}q_n + q_{n-1})},$$

for appropriate $a_{n+1} \leq \alpha_{n+1} < a_{n+1} + 1$. Recall that $q_n \geq q_{n-1}$. Thus

$$a_{n+1}q_n > \alpha_{n+1}q_n + q_{n-1} < (a_{n+1} + 2)q_n,$$

so we are done. □

Remark 5. If $0 < q < q_{n+1}$, then $|q\alpha - p| \geq |q_n\alpha - p_n|$. Also,

$$\det \begin{bmatrix} p_n & p_{n+1} \\ q_n & q_{n+1} \end{bmatrix} = (-1)^{n+1},$$

so there exist integers u, v such that

$$p = up_n + vp_{n+1}, q = uq_n + vq_{n+1},$$

with $u \geq 0$ and u, v having different signs. Hence,

$$\begin{aligned} |q\alpha - p| &= |\alpha(uq_n + vq_{n+1}) - (up_n + vp_{n+1})| \\ &= |u(\alpha q_n - p_n) + v(\alpha q_{n+1} - p_{n+1})| \\ &\geq |u + v||q_n\alpha - p_n| \geq |q_n\alpha p_n|. \end{aligned}$$

Proposition 3. *Let $p/q \in \mathbb{Q}$ and $\alpha \in \mathbb{R}$. If $\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}$, then $\frac{p}{q} = \frac{p_n}{q_n}$ (n-th convergent) for some n.*

Proof. For some n, we have $q_n \leq q < q_{n+1}$. Also,

$$\begin{aligned} \left|\frac{p}{q} - \frac{p_n}{q_n}\right| &\leq \left|\alpha - \frac{p}{q}\right| + \left|\alpha - \frac{p_n}{q_n}\right| \\ &\leq \frac{1}{q}|q\alpha - p| + \frac{1}{q_n}|q_n\alpha - p_n| \\ &\leq \left(\frac{1}{q} + \frac{1}{q_n}\right)|q\alpha - p| \\ &< \frac{2}{q_n} \cdot \frac{1}{2q} = \frac{1}{qq_n}. \end{aligned}$$

So if $\frac{p}{q} \neq \frac{p_n}{q_n}$, then $\left|\frac{p}{q} - \frac{p_n}{q_n}\right| \geq \frac{1}{qq_n}$, a contradiction. The claim follows. □

Definition 18. A continued fraction $[a_0, a_1, \ldots]$ is *ultimately periodic* if $a_{n+k} = a_n$ for all $n \geq m$ for some m and k. A continued fraction is *purely periodic* if $a_{n+k} = a_n$ for all $n \geq 0$ and some k.

Theorem 15 (Lagrange‘s theorem)**.** *A real number α is a quadratic irrational if and only if its continued fraction is ultimately periodic.*

Proof. ($\Rightarrow$) Suppose that $ax^2 + bx + c$ be the minimal polynomial of α. Note that $b^2 - 4ac > 0$. Thus we can let $\alpha = [a_0, a_1, \ldots, a_{n-1}, \alpha_n]$, or equivalently $\alpha = \frac{p_{n-1}\alpha_n + p_{n-2}}{q_{n-1}\alpha_n + q_{n-2}}$. Since $a\alpha^2 + b\alpha + c = 0$, there exist appropriate A_n, B_n, C_n such that $A_n\alpha_n^2 + B_n\alpha_n + C_n = 0$. Specifically,

$$\begin{aligned} A_n &= ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2 \\ B_n &= 2ap_{n-1}q_{n-2} + b(p_{n-1}q_{n-2} + p_{n-2}q_{n-1}) + 2cq_{n-1}q_{n-2} \\ C_n &= ap_{n-2}^2 + bp_{n-2}q_{n-2} + cq_{n-2}^2 = A_{n-1}. \end{aligned}$$

We will show that A_n, B_n, C_n are bounded. That is, there exist n_1, n_2, n_3 such that

$$\begin{aligned} (A, B, C) &= (A_{n_1}, B_{n_1}, C_{n_1}) \leftarrow \alpha_{n_1} \\ &= (A_{n_2}, B_{n_2}, C_{n_2}) \leftarrow \alpha_{n_2} \\ &= (A_{n_3}, B_{n_3}, C_{n_3}) \leftarrow \alpha_{n_3} \end{aligned}$$

At least two of $\alpha_{n_1}, \alpha_{n_2}, \alpha_{n_3}$ are the same. Without loss of generality, lets say $\alpha_{n_1} = \alpha_{n-2}(n_1 < n_2)$. Then $a_{n+k} = a_n$ for all $n \geq n_1$ and $k = n_2 - n_1$. Define $a_{n_1} := \lfloor \alpha_{n_1} \rfloor$ and $a_{n_2} := \lfloor \alpha_{n_2} \rfloor$. Since

$$\begin{aligned} \alpha_{n_1} &= a_{n_1} + \frac{1}{\alpha_{n_1+1}} \\ \alpha_{n_2} &= a_{n_2} + \frac{1}{\alpha_{n_2+1}}, \end{aligned}$$

we have $a_{n_1+1} = a_{n_2+1}$.

Note that $A_n \neq 0$: otherwise, $ax^2+bx+c=0$ has a rational root. Furthermore, $B_n^2 - 4A_nC_n = b^2-4ac > 0$. And if $\alpha - \frac{p_n}{q_n} = \frac{s_n}{q_n^2}$, then $|s_n| \le 1$. Substitution gives

$$\begin{aligned} A_n &= a\left(q_{n-1}\alpha - \frac{s_{n-1}}{q_{n-1}}\right)^2 + b\left(q_{n-1}\alpha - \frac{s_{n-1}}{q_{n-1}}\right)q_{n-1} + cq_{n-1}^2 \\ &= (a\alpha^2+b\alpha+c)q_{n-1}^2 - 2a\alpha s_{n-1} + a\frac{s_{n-1}^2}{q_{n-1}^2} - bs_{n-1} \\ &= -2a\alpha s_{n-1} + a\frac{s_{n-1}^2}{q_{n-1}^2} - bs_{n-1}. \end{aligned}$$

So $|A_n| \le |2a\alpha| + |a| + |b|$ and $|B_n| = |4A_nC_n + b^2 - 4ac| \le |4A_nC_n| + |b^2-4ac|$. Note $C_n = A_{n-1}$ so all three are bounded.

($\Leftarrow$) Write $\alpha = [a_0, a_1, \dots, a_{n-1}, \overline{a_n, a_{n+1}, \dots, a_{n+k-1}}]$, where the bar indicates periodicity. Let $\theta := [\overline{a_n, a_{n+1}, \dots, a_{n+k-1}}] \in \mathbb{R}\setminus\mathbb{Q}$. Let u_j/v_j be the j-th convergent to θ. Then $\theta = [a_n, a_{n+1}, \dots, a_{n+k-1}, \theta]$, or $\theta = \frac{u_{k-1}\theta + u_{k-2}}{v_{k-1}\theta+v_{k-2}}$. Hence $v_{k-1}\theta^2 + (v_{k-2}-u_{k-1})\theta - u_{k-2} = 0$. But since $\theta \notin \mathbb{Q}$, it follows that θ is a quadratic irrational. Now since $\alpha = [a_0, \dots, a_{n-1}, \theta]$, we have $\alpha = \frac{p_{n-1}\theta + p_{n-2}}{q_{n-1}\theta + q_{n-2}}$. α is a real quadratic rational as required. □

11. October 9

Proposition 4. *The continued fraction of α is purely periodic if and only if $\alpha > 1$ and its conjugate satisfies β satisfies $-1 < \beta < 0$.*

Proof. ($\Leftarrow$) We first prove this claim:

Claim. $-1 < \beta_n < 0$ where β_n is the conjugate of α_n.

We prove by induction on n. Suppose $a_n > 1$ and $-1 < \beta_n < 0$. Note that if

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

then

$$\beta_n = a_n + \frac{1}{\beta_{n+1}}.$$

Thus we have

$$\frac{1}{\beta_{n+1}} = \beta_n - a_n < -1.$$

so indeed $-1 < \beta_{n+1} < 0$, as desired.

Since

$$a_n = \beta_n - \frac{1}{\beta_{n+1}},$$

it follows

$$a_n = \left\lfloor -\frac{1}{\beta_{n+1}} \right\rfloor. \tag{18}$$

Since α is quadratic irrational, there exist $m, n \in \mathbb{Z}_+$ ($m > n$) such that $\alpha_m = \alpha_n$. In this case $\beta_m = \beta_n$. This implies $a_{n-1} = a_{m-1}$ (by (18)) so $\alpha_{n-1} = \alpha_{m-1}$. Repeating this argument yields $\alpha_0 = \alpha_{m-n}$. So the given continued fraction is purely periodic, as required.

($\Rightarrow$) Suppose that the continued fraction of α is purely periodic. Then $\alpha = [a_0, a_1, \ldots, a_n, \alpha]$ for some n. So there exist p_n and q_n such that

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

or equivalently, we have

$$q_n\alpha^2 + (q_{n-1} - p_n)\alpha - p_{n-1} = 0.$$

Let $f_n(x) := q_nx^2 + (q_{n-1} - p_n)x - p_{n-1}$. Then $f_n(0) = -p_{n-1} < 0$, and $f_n(-1) = q_n + p_n - q_{n-1} - p_{n-1} = (q_n - q_{n-1}) + (p_n - p_{n-1}) > 0$, since both $q_n - q - n - 1$ and $p_n - p_{n-1}$ are positive. Note that $p_n, q_n > 0$ and p_n, q_n are increasing sequences. Thus there exists a root $\beta \in (-1, 0)$, and so β is a conjugate of α as desired. □

Remark 6. Suppose that d is not a perfect square, and that

$$\alpha = \frac{1}{\sqrt{d} - \left\lfloor\sqrt{d}\right\rfloor} > 1.$$

Its conjugate is thus

$$\beta = \frac{1}{-\sqrt{d} - \left\lfloor\sqrt{d}\right\rfloor} = -\frac{1}{\sqrt{d} + \left\lfloor\sqrt{d}\right\rfloor}.$$

We have $-1 < \beta < 0$, so the continued fraction of α is purely periodic.

Consider the rational $\alpha = [a_0, \ldots, a_n]$, and let $\frac{p_i}{q_i}$ $(1 \le i \le n)$ be convergents of α. We state the following claims; we will only prove the first one.

Claim. We have

(1) $[a_n, a_{n-1}, a_{n-2}, \ldots, a_1, a_0] = \frac{p_n}{p_{n-1}}$
(2) $[a_n, a_{n-1}, \ldots, a_1] = \frac{q_n}{q_{n-1}}$.

Proof. We start from the fact that $p_n = a_np_{n-1} + p_{n-2}$. From which it follows

$$\begin{aligned}\frac{p_n}{p_{n-1}} &= a_n + \frac{1}{\frac{p_{n-1}}{p_{n-2}}}\\ &= a_n + \frac{1}{a_{n-1} + \frac{1}{\frac{p_{n-2}}{p_{n-3}}}}\\ &= \cdots = a_n + \frac{1}{a_{n-1} + \frac{1}{a_{n-2}} + \cdots + \frac{1}{\frac{p_1}{p_0}}},\end{aligned}$$

so the claim follows upon observing $\frac{p_1}{p_0} = a_0$. The proof of the second part follows in a similar manner. □

Proposition 5. *Let α be a quadratic irrational with $\alpha > 1$ and $-1 < \beta < 0$. Then we have*

$$\alpha = [\overline{a_0, a_1, \ldots, a_n}]$$
$$-\frac{1}{\beta} = [\overline{a_n, a_{n-1}, \ldots, a_1, a_0}].$$

Proof. Let $\theta = [\overline{a_n, a_{n-1}, \dots, a_1, a_0}] = [a_n, a_{n-1}, \dots, a_1, a_0, \theta]$, and let $\frac{u_n}{v_n}$ be the convergents to θ. That is, we have

$$\theta = \frac{u_n\theta + u_{n-1}}{v_n\theta + v_{n-1}}.$$

Let $\frac{p_n}{q_n}$ be the convergents of α. Note that by the first claim of the above remark, we have $\frac{u_n}{v_n} = \frac{p_n}{p_{n-1}}$. Since $(p_n, p_{n-1}) = (u_n, v_n) = 1$, we have $u_n = p_n$ and $v_n = p_{n-1}$. Also, by the second claim of the above remark, $\frac{u_{n-1}}{v_{n-1}} = \frac{q_n}{q_{n-1}}$. Thus $u_{n-1} = q_n$ and $v_{n-1} = q_{n-1}$. That is, we have

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

and equivalently

$$q_n\left(-\frac{1}{\theta}\right)^2 + (q_{n-1} - p_n)\left(-\frac{1}{\theta}\right) - p_{n-1} = 0.$$

Since α is also a root of $q_n x^2 + (q_{n-1} - p_n)x - p_{n-1} = 0$, indeed $-\frac{1}{\theta} = \beta$ is a conjugate of α. Hence $\theta = -\frac{1}{\beta}$. □

Claim. $\sqrt{d} = [a_0, \overline{a_1, \dots, a_n, 2a_0}]$.

Proof. Let $\alpha = \sqrt{d} + \left\lfloor \sqrt{d} \right\rfloor$. Then its conjugate β is $\beta = -\sqrt{d} + \left\lfloor \sqrt{d} \right\rfloor$. Since $-1 < \beta < 0$ and

$$\alpha = [\overline{2\left\lfloor \sqrt{d} \right\rfloor, a_1, a_2, \dots, a_n}] = [\overline{2a_0, a_1, \dots, a_n}],$$

By Prop 5, we get

$$-\frac{1}{\beta} = [\overline{a_n, a_{n-1}, \dots, a_1, 2a_0}] = \frac{1}{\sqrt{d} - \left\lfloor \sqrt{d} \right\rfloor}.$$

We have

$$\begin{aligned}\sqrt{d} - \left\lfloor \sqrt{d} \right\rfloor &= 0 + \frac{1}{\frac{1}{\sqrt{d} - \left\lfloor \sqrt{d} \right\rfloor}} = 0 + \frac{1}{[\overline{a_n, a_{n-1}, \dots, a_1, 2a_0}]} \\ &= [0, \overline{a_n, a_{n-1}, \dots, a_1, 2a_0}].\end{aligned}$$

Thus

$$\begin{aligned}\sqrt{d} &= \underbrace{\left\lfloor \sqrt{d} \right\rfloor}_{=a_0} + [0, \overline{a_n, a_{n-1}, \dots, a_1, 2a_0}] \\ &= [a_0, \overline{a_n, a_{n-1}, \dots, a_2, a_1, 2a_0}] \\ \alpha = \sqrt{d} + \left\lfloor \sqrt{d} \right\rfloor &= [\overline{2a_0, a_1, a_2, \dots, a_n}].\end{aligned}$$

Recall that $2a_0 = 2\left\lfloor\sqrt{d}\right\rfloor$, so

$$\begin{aligned}\sqrt{d} &= -\left\lfloor\sqrt{d}\right\rfloor + [\overline{2a_0, a_1, a_2, \dots, a_n}] \\ &= a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_n + \frac{1}{2a_0}}}} \\ &= [a_0, \overline{a_1, a_2, \dots, 2a_0}].\end{aligned}$$

Note that this gives us $\sqrt{d} = [a_0, \overline{a_1, \dots, a_n, 2a_0}] = [a_0, \overline{a_n, a_{n-1}, \dots, a_1, 2a_0}]$, so $a_1 = a_n, a_2 = a_{n-1}, \dots$. Hence $\sqrt{d} = [a_0, \overline{a_1, a_2, \dots, a_2, a_1, 2a_0}]$. □

12. October 14

Let $d \in \mathbb{N}, d > 1$, not a square. The equation

$$x^2 - dy^2 = 1$$

in integers x and y is known as the Pell equation. Fermat conjectured that the equation always has a non-trivial solution i.e., different from $(x, y) = (\pm 1, 0)$. This was first proved by Lagrange in 1768.

Let us consider the equations

$$x^2 - dy^2 = 1 \tag{19}$$

$$x^2 - dy^2 = -1. \tag{20}$$

Suppose that (x, y) is a nontrivial positive solution to (19) and (20). Then

$$x \geq \sqrt{dy^2 - 1} \geq y\sqrt{d - 1}.$$

Thus we have

$$|x - \sqrt{d}y| = \frac{1}{|x + \sqrt{d}y|} \leq \frac{1}{\sqrt{d}y\left(1 + \sqrt{1 - \frac{1}{d}}\right)} < \frac{1}{2y},$$

since $\sqrt{d} + \sqrt{d - 1} > 2$ for $d \geq 2$. Thus

$$\left|\sqrt{d} - \frac{x}{y}\right| < \frac{1}{2y^2}.$$

Therefore $x | y$ is a convergent to $\sqrt{d}$, i.e., $\frac{x}{y} = \frac{p_n}{q_n}$ for some $n \geq 0$. Then

$$\sqrt{d} = \frac{p_n \alpha_{n+1} + p_{n-1}}{q_n \alpha_{n+1} + q_{n-1}}.$$

so

$$\begin{aligned}(p_n - q_n\sqrt{d})\alpha_{n+1} &= \sqrt{d}q_{n-1} - p_{n-1} \\ (p_n^2 - q_n^2 d)\alpha_{n+1} &= (\sqrt{d}q_{n-1} - p_{n-1})(p_n + q_n\sqrt{d}) \qquad (21) \\ (\pm 1)\alpha_{n+1} &= \sqrt{d}(p_n q_{n-1} - p_{n-1}q_n) - p_{n-1}p_n + q_{n-1}q_n d \\ &= \sqrt{d}(-1)^{n+1} + h \text{ for } h \in \mathbb{Z}.\end{aligned}$$

The convergents of even index are smaller than $\sqrt{d}$ and the convergents of odd index are larger than $\sqrt{d}$. Therefore, by (21), we have that $n-1$ is even if $p_n^2 - dq_n^2 = 1$, and $n-1$ is odd if $p_n^2 - dq_n^2 = -1$.

Consider the first possibility. Then

$$\alpha_{n+1} = \sqrt{d}(-1)^{n+1} + h = \sqrt{d} + h.$$

Thus $\alpha_{n+2} = \alpha_1$. But $\sqrt{d} = [a_0, \overline{a_1, a_2, \ldots, a_m}]$ where m is the minimal length of the period of the continued-fraction expansion of $\sqrt{d}$. We have then $\alpha_1 = \alpha_{m+1} = \alpha_{2m+1} = \cdots$, and $\alpha_k \neq \alpha_1$ for $k \not\equiv 1 \pmod{m}$ since m is the minimal period. Thus $m \mid (n+2) - 1$ hence $n = lm - 1$ for some $l \in \mathbb{Z}_+$. In this case we have $n-1$ even so lm is even. In the case when $x^2 - dy^2 = -1$ we find that lm is odd so

$$-\alpha_{n+1} = -\sqrt{d} + h, \text{ or } \alpha_{n+1} = \sqrt{d} - h.$$

In particular, if the minimal period m is even, the equation has $x^2 - dy^2 = -1$ has no non-trivial solution.

Theorem 16. *Let d be a squarefree integer with $d > 1$. Let m be the minimal period of the continue-fraction (CF) expansion of $\sqrt{d}$. (x, y) is the solution of*

$$x^2 - dy^2 = 1$$

with $x, y \in \mathbb{N}$ iff $x = p_n, y = q_n$ for some convergent p_n/q_n to $\sqrt{d}$ and where $n = lm - 1, l \geq 0, lm$ even. Also, (x, y) is a solution of

$$x^2 - dy^2 = -1$$

with $x, y \in \mathbb{N}$ iff $x = p_n, y - q_n, lm$ is odd, $l > 0, n = lm - 1$.

Proof. The above discussion already established ($\Rightarrow$).

($\Leftarrow$) Suppose that $n = lm - 1$. Then by periodicity $\alpha_1 = \alpha_{n+2}$ so

$$\sqrt{d} = \frac{p_{n+1}\alpha_{n+2} + p_n}{q_{n+1}\alpha_{n+2} + q_n} = \frac{p_{n+1}\alpha_1 + p_n}{q_{n+1}\alpha_1 + q_n}.$$

But $\alpha_1 = \frac{1}{\sqrt{d} - a_0}$ and so

$$(q_{n+1} + q_n(\sqrt{d} - a_0))\sqrt{d} = p_{n+1} + p_n(\sqrt{d} - a_0).$$

From this, we have $q_{n+1} - q_n a_0 = p_n$ and $q_n d = p_{n+1} - p_n a_0$ since $\sqrt{d} \notin \mathbb{Q}$. Eliminating a_0 gives us $p_n q_{n+1} - p_{n+1} q_n = p_n^2 - q_n^2 d$, hence $p_n^2 - q_n^2 d = (-1)^{n+1}$. Thus if n is odd then we have a non-trivial solution of $x^2 - dy^2 = 1$ while if n is even we have a non-trivial solution of $x^2 - dy^2 = -1$. □

13. October 16 & 19

In general, not much is known about the continued fraction expansion of algebraic numbers of degree greater than 2. No such number is known to have bounded partial quotients. For certain non-algebraic numbers of interest we know more. For example,

$$e - 1 = [1, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots].$$

As a consequence, we are able to show that there is a positive number c such that for $q > 4$, we have

$$\left|e - \frac{p}{q}\right| > \frac{c \log\log q}{q^2 \log q}.$$

On the other hand, π is a mystery:

$$\pi = [3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, \ldots].$$

The initial convergents are $3, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103993}{33102}, \ldots$. We have

$$\left|\pi - \frac{22}{7}\right| = \frac{1}{(16.139\ldots)7^2}, \left|\pi - \frac{355}{113}\right| = \frac{1}{(293.57\ldots)(113)^2}.$$

In fact, $\frac{22}{7}$ is as good as $\frac{p}{q}$ for any $q < 57$; similarly, $\frac{335}{113}$ is as good as $\frac{p}{q}$ for $q < 16604$.

Mahler in 1953 proved that there exists a positive number c such that

$$\left|\pi - \frac{p}{q}\right| > \frac{c}{q^{42}}.$$

Also, there is a theorem by Salikhov

Theorem 17 (Salikhov, 2008). $\left|\pi - \frac{p}{q}\right| > \frac{1}{q^{7.6063\ldots}}$ *for q sufficiently large.*

What does the continued fraction expansion of a "typical" real number look like? A first question one might ask is: how does q_n grow? For all irrational numbers α it grows exponentially. To see this observe that $q_0 = 1$ and $q_1 = a_1$ and for $n \geq 2$ we have $q_n = a_n q_{n-1} + q_{n-2}$. Thus $q_n \geq u_{n+1}$ for $n = 0, 1, 2, \ldots$ where $u_0 = 0, u_1 = 1$ and $u_n = u_{n-1} + u_{n-2}$ for $n \geq 2$. But $(u_n)_{n=0}^{\infty}$ is the Fibonacci sequence and

$$u_n = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}}.$$

Theorem 18. *There exists a positive number c such that for all real numbers α, except a set of Lebesgue measure zero, we have*

$$q_n = q_n(\alpha) < e^{cn}.$$

Proof (Khintchine). We first remark that we can restrict our attention to α in $(0, 1)$ since the countable union of sets of measure zero is a set of measure zero. Let $E_n(g)$ for $n \geq 1, g \geq 1$ be the set of real numbers in $(0, 1)$ for which $a_1 \ldots a_n \geq g$ where $a_1, \ldots, a_n$ are the initial partial quotients of α.

For any fixed sequence $(a_1, \ldots, a_n)$ we will determine the measure of the set of α's in $(0, 1)$ whose first $n+1$ partial quotients are $0, a_1, \ldots, a_n$. Thus

$$\alpha = [0, a_1, \ldots, a_n, \alpha_{n+1}].$$

Then

$$\frac{p_n \alpha_{n+1} + p_{n-1}}{q_n \alpha_{n+1} + q_{n-1}}$$

and α_{n+1} varies from 1 to ∞. This gives an interval with endpoints

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

observe that

$$\alpha - \frac{p_n}{q_n} = \frac{p_n\alpha_{n+1} + p_{n-1}}{q_n\alpha_{n+1} + q_{n-1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(q_n\alpha_{n+1} + q_{n-1})}$$

is a monotone function of α_{n+1}. The length of the interval is

$$\left|\frac{p_n + p_{n-1}}{q_n + q_{n-1}} - \frac{p_n}{q_n}\right| = \frac{1}{q_n(q_n + q_{n-1})}.$$

and

$$\frac{1}{q_n(q_n + q_{n-1})} < \frac{1}{q_n^2} < \frac{1}{(a_1 \dots a_n)^2}$$

since $q_n > a_n q_{n-1}$ hence $q_n > a_n \cdots a_1$. Thus the measure

$$\mu(E_n(g)) < \sum_{a_1 \dots a_n \geq g} \frac{1}{(a_1 \dots a_n)^2}.$$

Note that

$$\begin{aligned}\prod_{i=1}^{n} \frac{1}{a_i^2} &= \prod_{i=1}^{n}\left(1 + \frac{1}{a_i}\right) \cdot \frac{1}{a_i(a_i+1)} \\ &\leq 2^n \prod_{i=1}^{n} \frac{1}{a_i(a_i+1)} \\ &= 2^n \prod_{i=1}^{n} \int_{a_i}^{a_i+1} \frac{dx_i}{x_i^2} \\ &= 2^n \int_{a_1}^{a_1+1} \cdots \int_{a_n}^{a_n+1} \frac{dx_1 dx_2 \dots dx_n}{x_1^2 \dots x_n^2}.\end{aligned}$$

Put

$$J_n(g) = \int_R \frac{dx_1 \dots dx_n}{x_1^2 \dots x_n^2}$$

where R is the region $x_i \geq 1$ for $i = 1, \dots, n$ and $x_1 \cdots x_n \geq g$. Thus

$$\mu(E_n(g)) < 2^n J_n(g).$$

It remains to evaluate $J_n(g)$. If $g \leq 1$ then

$$R = \{(x_1, \dots, x_n) : x_i \geq 1 \text{ for } i = 1, \dots, n\}$$

and so

$$J_n(g) = \left(\int_1^\infty \frac{dx}{x^2}\right)^n = 1.$$

We now prove by induction on n that for $g > 1$,

$$J_n(g) = \frac{1}{g}\sum_{i=0}^{n-1} \frac{(\log g)^i}{i!}. \tag{22}$$

For $n = 1$ we have $J_1(g) = \int_g^\infty \frac{dx}{x^2} = \frac{1}{g}$ as expected. Now assume the result for $n = k$. Then

$$J_{n+1}(g) = \int_1^\infty \frac{dx_{k+1}}{x_{k+1}^2} J_k\left(\frac{g}{x_{k+1}}\right).$$

Apply the change of variables

$$u = \frac{g}{x_{k+1}} \text{ so } du = -\frac{g}{x_{k+1}^2} dx_{k+1}.$$

Thus

$$\begin{aligned} J_{k+1}(g) &= \frac{1}{g}\int_0^g J_k(u)\,du \\ &= \frac{1}{g}\left(\int_0^1 J_k(u)\,du + \int_1^g J_(u)\,du\right) \\ &= \frac{1}{g}\left(1 + \int_1^g \frac{1}{u}\sum_{i=0}^{k-1}\frac{(\log u)^i}{i!}\,du\right) \\ &= \frac{1}{g}\left(1 + \sum_{i=0}^{k-1}\frac{(\log u)^{i+1}}{(i+1)!}\Bigg|_1^g\right) \\ &= \frac{1}{g}\left(1 + \sum_{i=0}^{k-1}\frac{(\log g)^{i+1}}{(i+1)!}\right) = \frac{1}{g}\sum_{i=0}^{k}\frac{(\log g)^i}{i!}. \end{aligned}$$

This completes the induction argument. We now put $g = e^{An}$ where $A > 1$ is to be chosen. We now find that

$$\mu(E_n(e^{An})) < \frac{2^n}{e^{An}}\sum_{i=0}^{n-1}\frac{(An)^i}{i!} < \frac{2^n A^n}{e^{An}}\sum_{i=0}^{n-1}\frac{n^i}{i!} < \frac{2^n A^n}{e^{An}}e^n = e^{(1+\log 2+\log A - A)n}.$$

Now choose A so that $1 + \log 2 + \log A - A < 0$, from which it follows

$$\sum_{n=1}^{\infty}\mu_n(e^{An}) < \infty.$$

Thus, by Borel-Cantelli, every number α, apart from a set of measure zero, belongs to only finitely many of the sets $E_n(e^{An})$. Thus for almost all $\alpha \in (0,1)$, there exists $N(\alpha)$ such that for $n > N(\alpha)$, we have $a_1 \dots a_n < e^{An}$. But $q_n = a_n q_{n-1} + q_{n-2} < 2a_n q_{n-1}$, hence $q_n < a_1 a_2 \dots a_n$. Thus for almost all α and for n sufficiently large in terms of α, we indeed have $q_n < e^{(A+\log 2)n}$ as required. □

In 1935, Paul Lévy proved that for almost all α in the sense of Lebesgue measure we have

$$\lim_{n\to\infty}(q_n(\alpha))^{1/n} = e^{\frac{\pi^2}{12\log 2}}.$$

He used probability theory. We will prove this using ergodic theory instead. Consider the probability space $(X, \mathcal{B}, \mu)$ consisting of a space X, a σ-algebra $\mathcal{B}$ and a non-negative, countably additive measure μ on X with $\mu(X) = 1$. We say that T is a measure-preserving transformation of $(X, \mathcal{B}, \mu)$ if $T : X \to X$, $\mu(T^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}$, and $B \in \mathcal{B} \Rightarrow T^{-1}(B) \in \mathcal{B}$. Let $\mathcal{L}'$ consist of the measurable functions $f : X \to \mathbb{R}$ which are integrable. Then if T is measure-preserving and $f \in \mathcal{L}'$ then $\int f\,d\mu = \int f \circ T\,d\mu$.

Definition 19. Let T be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. Then T is said to be *ergodic* if whenever $B \in \mathcal{B}$ and $T^{-1}B \subseteq B$ then $\mu(B) = 0$ or $\mu(B) = 1$.

Theorem 19 (Ergodic theorem)**.** *Suppose $f \in \mathcal{L}'$ and T is ergodic. Then, for almost all $x \in X$, we have*

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

Proof. The proof can be found in a book on ergodic theory. □

14. October 21

Let $X = (0,1)$. Let $\mathcal{B}$ be the σ-algebra of Lebesgue measurable sets on $(0,1)$ and let μ be the Lebesgue measure on $(0,1)$. Let $T : X \to X$ by $T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$. Note that T is *not* measure-preserving with respect to μ and so we modify μ. We define μ_1 to be the measure for which for all $f \in \mathcal{L}'$,

$$\mu_1(f) := \frac{1}{\log 2} \int_0^1 \frac{f(x)}{1+x} \, dx. \tag{23}$$

Observe that X is a probability space with respect to μ_1 since

$$\mu_1(\mathbf{1}_X) = 1.$$

Note that (23) determines μ_1. We claim that T is measure-preserving with respect to μ_1. instead of all measurable sets on X, it suffices to check that T is measure-preserving with respect to each interval (a,b). We have

$$T^{-1}(a,b) = \bigcup_{n=1}^{\infty} \left(\frac{1}{b+n}, \frac{1}{a+n} \right),$$

since if x is in $\left(\frac{1}{b+n}, \frac{1}{a+n}\right)$ then Tx is in (a,b). In particular, $T(1/(b+n), 1/(a+n)) = (a,b)$ for all $n \in \mathbb{Z}_+$. Further, there is no longer set sent to (a,b) by T. Certainly, the set

$$\bigcup_{n=1}^{\infty} \left(\frac{1}{b+n}, \frac{1}{a+n} \right)$$

is measurable. Further,

$$\begin{aligned}
\mu_1(T^{-1}(a,b)) = \sum_{n=1}^{\infty} \mu_1 \left(\frac{1}{b+n}, \frac{1}{a+n} \right) &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{\frac{1}{b+n}}^{\frac{1}{a+n}} \frac{dx}{1+x} \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \log(1+x) \Big|_{\frac{1}{b+n}}^{\frac{1}{a+n}} \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log\left(1 + \frac{1}{a+n}\right) - \log\left(1 + \frac{1}{b+n}\right) \right) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log\left(\frac{a+n+1}{a+n}\right) - \log\left(\frac{b+n+1}{b+n}\right) \right) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log\left(\frac{a+n+1}{b+n+1}\right) - \log\left(\frac{a+n}{b+n}\right) \right).
\end{aligned}$$

But

$$\sum_{n=1}^{N} \log\left(\frac{a+n+1}{a+n}\right) - \log\left(\frac{b+n+1}{b+n}\right) = \sum_{n=1}^{N} \log\left(\frac{a+n+1}{b+n+1}\right) - \log\left(\frac{a+n}{b+n}\right)$$

$$= \log\left(\frac{a+N+1}{b+N+1}\right) - \log\left(\frac{a+1}{b+1}\right);$$

and as $N \to \infty$, this tends to $\log\left(\frac{b+1}{a+1}\right)$. Therefore

$$\mu_1(T^{-1}(a,b)) = \frac{1}{\log 2}\log\left(\frac{b+1}{a+1}\right) = \frac{1}{\log 2}\int_a^b \frac{dx}{1+x} = \mu_1(a,b).$$

Thus μ_1 is invariant with respect to T. This was understood by Gauss in 1812.

Note that $\alpha_n = a_n + \frac{1}{\alpha_{n+1}}$, for $n = 0, 1, 2, \ldots$. Further $\alpha_i \geq 1$ for $i = 1, 2, \ldots$,

$$T\left(\frac{1}{\alpha_n}\right) = \alpha_n - \lfloor \alpha_n \rfloor = \alpha_n - a_n = \frac{1}{\alpha_{n+1}}$$

for $n = 1, 2, \ldots$. It can be proved that T is ergodic. Thus we can take k to be a positive integer and f to be the characteristic function of $((k+1)^{-1}, k^{-1})$. We now apply the Ergodic theorem to deduce that for almost all x in the sense of Lebesgue measure,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) = \int_X f\, d\mu_1 = \frac{1}{\log 2}\int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{dx}{1+x}$$

$$= \frac{1}{\log 2}\left(\log\left(1+\frac{1}{k}\right) - \log\left(1+\frac{1}{k+1}\right)\right) = \frac{1}{\log 2}\log\left(\frac{(k+1)^2}{k(k+2)}\right).$$

This tells us that for "almost all" real numbers α the frequency with which the n-th partial quotient of α is k exists and is equal to

$$\frac{1}{\log 2}\log\left(\frac{(k+1)^2}{k(k+2)}\right).$$

Gauss conjectured this fact, and it was first proved by Kuzmin in the 1920's. The Gauss-Kuzmin theorem tells us that for almost all α in the sense of Lebesgue measure the frequency of 1's is .41503 . . . , of 2's is .169925 . . . , 3's is .0931 . . . , 4's is .0588 . . . , and 5's .0406 . . . . The expected frequency of odd partial quotients is

$$\theta = \frac{1}{\log 2}\sum_{j=1}^{\infty} \log\left(\frac{(2j)^2}{(2j-1)(2j+1)}\right)$$

and the frequency of even partial quotients for almost all α is $1-\theta$. But $\theta > \frac{1}{2}$ and this contrasts with the fact that for almost all real numbers α in the sense of Lebesgue measure the frequency of even decimal digits in the base-10 expansion of α is $\frac{1}{2}$ as is the frequency of odd decimal digits. The frequency with which a partial quotient of α is at least k is

$$\frac{1}{\log 2}\sum_{j=k}^{\infty} \log\left(\frac{(j+1)^2}{j(j+2)}\right)$$

for all real numbers α except on a set of measure zero.

15. October 23

Observe that if $\alpha = \alpha_0 \in (0,1)$ then $\alpha_0 = \alpha_1^{-1}$. Thus

$$T(\alpha) = \frac{1}{\alpha_0} - \left\lfloor \frac{1}{\alpha_0} \right\rfloor = \alpha_1 - \lfloor \alpha_1 \rfloor = \frac{1}{\alpha_2}.$$

So $T(\alpha) = T(\alpha^{-1}) = \alpha_2^{-1}$. More generally, we have $T^n(\alpha) = T^n(\alpha_1^{-1}) = \alpha_{n+1}^{-1}$ for $n \geq 0$. Further, let $a_n = \lfloor \alpha_n \rfloor$ for $n \geq 0$. Thus

$$\sqrt[n]{a_1 a_2 \cdots a_n} = \sqrt[n]{\lfloor T^0(\alpha)^{-1} \rfloor \lfloor T_1(\alpha)^{-1} \rfloor \cdots \lfloor T^{n-1}(\alpha)^{-1} \rfloor},$$

and so

$$\frac{1}{n} \sum_{j=0}^{n-1} \log a_j = \frac{1}{n} \sum_{j=0}^{n-1} \log \left(\left\lfloor \frac{1}{T^j(\alpha)} \right\rfloor \right).$$

Take $f(x) = \log \lfloor x^{-1} \rfloor$ and apply the Ergodic theorem to deduce that for almost all x in $(0,1)$ in the sense of Lebesgue measure we have

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \log a_i &= \frac{1}{\log 2} \int_0^1 \log \left\lfloor \frac{1}{x} \right\rfloor \frac{dx}{x+1} \\ &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{\frac{1}{n+1}}^{\frac{1}{n}} \log n \frac{dx}{x+1} \\ &= \sum_{n=1}^{\infty} \frac{\log n}{\log 2} \log(1+x) \Big|_{\frac{1}{n+1}}^{\frac{1}{n}} \\ &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \log n \log \left(\frac{(n+1)^2}{n(n+2)} \right), \end{aligned}$$

or equivalently,

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

First note that for α in $(0,1)$ with the first $n+1$ partial quotients $0, a_1, \ldots a_n$ we have $[0, a_1, \ldots, a_n] = \frac{p_n}{q_n}$ so $[a_1, \ldots, a_n] = \frac{q_n}{p_n}$. We now note that

$$q_n = [a_1, \ldots, a_n][a_2, \ldots, a_n] \cdots [a_n]$$

since if $[a_j, \ldots, a_n] = \frac{x}{b}$ then $[a_{j+1}, \ldots, a_n] = \frac{b}{c}$. So we have a telescoping product with first term $\frac{q_n}{p_n}$ and last term $\frac{a_n}{1}$. (Aside: since $q_j/q_{j-1} = [a_j, \ldots, a_1]$ for $j \geq 1$ and so we have $q_n = [a_n, \ldots, a_1][a_{n-1}, \ldots, a_1] \cdots [a_1]$.) So it follows

$$q_n = \left(\frac{p_n}{q_n} \right)^{-1} \left(T \left(\frac{p_n}{q_n} \right)^{-1} \right) \cdots \left(T^{n-1} \left(\frac{p_n}{q_n} \right)^{-1} \right).$$

We will now prove that if the first $n+1$ partial quotients of α are $0, a_1, \ldots, a_n$ then

$$\left| \log T^i(\alpha) - \log T^i \left(\frac{p_n}{q_n} \right) \right| < 2^{-\frac{1}{2}(n-i-1)+1}. \tag{24}$$

it suffices to prove (24) for $i = 0$ by induction on n. Since α is in an interval with endpoints $\frac{p_n}{q_n}$ and $\frac{p_n+p_{n-1}}{q_n+q_{n-1}}$. Hence

$$\left|\log\left(\frac{\alpha}{p_n/q_n}\right)\right| \leq \left|\log\left(\frac{\frac{p_n+p_{n-1}}{q_n+q_{n-1}}}{\frac{p_n}{q_n}}\right)\right|.$$

But

$$\left|\frac{q_n}{p_n}\cdot\frac{p_n+p_{n-1}}{q_n+q_{n-1}} - \frac{p_n(q_n+q_{n-1})}{p_n(q_n+q_{n-1})}\right| = \frac{1}{p_n(q_n+q_{n-1})}.$$

Thus $\log\left(\frac{\alpha}{p_n/q_n}\right) = \log(1+t)$ where $|t| < \frac{1}{p_n(q_n+q_{n-1})}$. Now $|\log(1-x)| < 2x$ and $|\log(1+x)| < x$ for $0 < x \leq \frac{1}{2}$. Thus

$$\left|\log\alpha - \log\frac{p_n}{q_n}\right| < \frac{2}{p_n(q_n+q_{n-1})}$$

for $n = 1, 2, \ldots$. Since $q_n \geq 2^{\frac{1}{2}(n-1)}$, it follows

$$\left|\log\alpha - \log\frac{p_n}{q_n}\right| < \frac{2}{2^{\frac{n-1}{2}}}$$

for $n = 1, 2, \ldots$, as required. This proves (24). Therefore

$$\begin{aligned}\left|\sum_{i=0}^{n}\left(\log T^i(\alpha) - \log T^i\left(\frac{p_n}{q_n}\right)\right)\right| &< 2\sum_{i=0}^{n-1} 2^{-\frac{1}{2}(n-i-1)} \\ &\leq 2\sum_{j=0}^{\infty}\left(\frac{1}{\sqrt{2}}\right)^j = \frac{2}{1-\frac{1}{\sqrt{2}}} = \frac{2\sqrt{2}}{\sqrt{2}-1} = 6.82\cdots < 7.\end{aligned}$$

Since

$$-\log q_n = \sum_{i=0}^{n-1}\log T^i\left(\frac{p_n}{q_n}\right),$$

we have

$$\left|\frac{1}{n}\left(sum_{i=0}^{n-1}\log T^i(\alpha) + \log q_n\right)\right| < \frac{7}{n},$$

and so for all irrational α we have

$$\lim_{n\to\infty}\frac{1}{n}\left(\sum_{i=0}^{n-1}\log(T^i(\alpha))^{-1} - \log q_n\right) = 0.$$

Thus by the Ergodic theorem with $f(x) = \log\frac{1}{x}$ for almost all x in the sense of Lebesgue measure it follows

$$\lim_{n\to\infty}\frac{1}{n}\log q_n = \lim_{n\to\infty}\sum_{i=0}^{n-1}\log(T^i\alpha)^{-1} = \frac{1}{\log 2}\int_0^1 \log\left(\frac{1}{x}\right)\frac{dx}{x+1}.$$

Equivalently,

$$\lim_{n\to\infty} q_n^{1/n} = e^{\frac{1}{\log 2}\int_0^1 \log(x^{-1})\frac{dx}{x+1}}$$

for all α except at a set of measure zero. To obtain Lévy's theorem, it suffices to prove that

$$\int_0^1 \log\left(\frac{1}{x}\right)\frac{dx}{x+1} = \frac{\pi^2}{12}.$$

And we shall continue next Monday.

16. October 26

Let $f(X) = \log x$ so $f'(x) = x^{-1}$. Let $g(x) = \log(1+x)$ hence $g'(x) = (x+1)^{-1}$. Thus we have

$$\int_0^1 (f(x)g'(x) + g(x)f'(x))\, dx = g(x)f(x)|_0^1 .$$

Thus it follows that

$$\int_0^1 \left(\frac{\log x}{x+1} + \frac{\log(1+x)}{x} \right) dx = \lim_{x \to 1} \log x \log(1+x) - \lim_{x \to 0} \log x \log(1+x) = 0 - 0 = 0.$$

Hence we have

$$\begin{aligned} \int_0^1 \log\left(\frac{1}{x}\right) \frac{dx}{x+1} &= \int_0^1 \log(1+x)\, \frac{dx}{x} = \int_0^1 \left(\sum_{h=1}^{\infty} \frac{(-1)^{h-1} x^h}{h} \right) \frac{dx}{x} \\ &= \int_0^1 \sum_{h=0}^{\infty} \frac{(-1)^h x^h}{h+1}\, dx = \sum_{h=0}^{\infty} \int_0^1 \frac{(-1)^h x^h}{h+1}\, dx \\ &= \sum_{h=0}^{\infty} \frac{(-1)^h}{(h+1)^2} = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \cdots \\ &= \sum_{h=1}^{\infty} \frac{1}{h^2} - 2 \sum_{h=1}^{\infty} \frac{1}{(2h)^2} = \frac{\pi^2}{12}, \end{aligned}$$

as we wanted.

Dobrowolski in 1979 proved that if $\varepsilon > 0$ and α is a non-zero algebraic number of degree d with

$$\mu(\alpha) < 1 + (1+\varepsilon) \left(\frac{\log \log d}{\log d} \right)^3$$

then for d sufficiently large in terms of ε, the number α is a root of unity. He needed three new ingredients. The first is a sharper version of Siegel's lemma. The proof is essentially along the same lines. We need the estimate for the size of the coefficients to improve if the number of variables is more than $2d$ times the number of unknowns. This feature goes back to Siegel.

Theorem 20 (Siegel's lemma II)**.** *Let $b_{ij}, 1 \le i \le N, 1 \le j \le M$, be algebraic integers in a field K such that for each J not all of the b_{ij}'s zero. Let $[K : \mathbb{Q}] := d$ and let $\sigma_1, \ldots, \sigma_d$ be the embeddings of K in $\mathbb{C}$. If $N > dM$ then the system of equations*

$$\sum_{j=1}^{N} b_{ij} x_i = 0, 1 \le j \le M$$

has a solution in rational integers $x_1, \ldots, x_N$, not all 0, whose absolute values are at most

$$\left(2\sqrt{2}(N+1) \left(\prod_{j=1}^{M} \prod_{k=1}^{d} \max_i |\sigma_k(b_{ij})| \right)^{\frac{1}{dM}} \right)^{\frac{dM}{N-dM}} .$$

Proof. Similar to the previous version! □

The sharpest form of Siegel's lemma known is due to Bombieri and Vaaler.

Lemma 2. *If α is an algebraic number of degree d and $P = \{p \text{ prime} : \deg(\alpha^p) < d\}$ then*

$$|P| \leq \frac{\log d}{\log 2}.$$

Proof. Let $\alpha_1, \ldots, \alpha_d$ be the conjugates of α. For integers s and j with $1 \leq j \leq d$ we put $I(s, j) := \{i : \alpha_i^s = \alpha_j^s\}$. We first note that

$$|I(s, j)| = |I(s, t)| \tag{25}$$

for $1 \leq t \leq d, 1 \leq j \leq d$, since there is an element $\lambda \in \mathrm{Gal}(\mathbb{Q}(\alpha_1, \ldots, \alpha_d)/\mathbb{Q})$ which sends α_j to α_t and induces a permutation on the remaining roots. This remark also tells us that if $I(s, j) \neq I(s, t)$ then

$$I(s, j) \cap I(s, t) = \emptyset. \tag{26}$$

Next we prove that if r and s are coprime then

$$|I(r, i) \cap I(s, j)| \leq 1. \tag{27}$$

To see this, suppose that $k, l \in I(r, i) \cap I(s, j)$. Then

$$(\alpha_k^s = \alpha_j^s) \wedge (\alpha_l^s = \alpha_j^s) \Rightarrow \alpha_k^s = \alpha_l^s$$

and

$$(\alpha_k^r = \alpha_j^r) \wedge (\alpha_l^r = \alpha_j^r) \Rightarrow \alpha_k^r = \alpha_l^r.$$

Therefore

$$\alpha_k^{(r,s)} = \alpha_l^{(r,s)} \Rightarrow \alpha_k = \alpha_l \Rightarrow k = 1,$$

as required. Next we observe that if $(r, s) = 1$ then

$$|I(rs, j)| \geq |I(r, j)| \cdot |I(s, j)|. \tag{28}$$

By (25) each of the $I(s, k)$'s have the same cardinality and by (26), we see that (28) follows. But then since

$$2^{|P|} \leq \prod_{p \in P} |I(p, i)| \leq \left| I\left(\prod_{p \in P} p, i\right)\right| \leq d,$$

it follows that

$$|P| \leq \frac{\log d}{\log 2},$$

as desired. □

17. October 28 & October 30

Remark 7. In the proof of Lemma 2, we needed coprimality to ensure every term in the union occurs at most one time.

The next lemma that we need is the crucial new ingredient in Dobrowolski's argument. It allows us to replace the lower bound of 1 for the absolute value of the norm of an algebraic integer with something much longer based on a congruence argument.

Lemma 3. *Let α be a non-zero algebraic integer of degree d with conjugates $\alpha = \alpha_1, \dots, \alpha_d$. Let f be the minimal polynomial of α over the integers. Suppose that α is not a root of unity. Then for integers r and s with $1 \leq s < r$, we have*

$$\alpha_i^r \neq \alpha_j^s \text{ for } 1 \leq i, j \leq d.$$

Further, we have

$$\left| \prod_{i=1}^{d} f(\alpha_i^p) \right| \geq p^d.$$

Proof. If $\alpha_i^r = \alpha_j^s$ then α_i^s is a conjugate of α_i^r. Thus there exists an element $\sigma \in \mathrm{Gal}(\mathbb{Q}(\alpha_1, \dots, \alpha_n)/\mathbb{Q})$ such that $\sigma(\alpha_i^r) = \alpha_i^s$. Let k be the order of γ in $\mathrm{Gal}(\mathbb{Q}(\alpha_1, \dots, \alpha_n)/\mathbb{Q})$. Thus $\sigma^k = \mathrm{id}$. Then

$$\begin{aligned} \alpha_i^{r^k} &= \sigma^k(\alpha_i^{r^k}) = (\sigma^k(\alpha_i^r))^{r^{k-1}} \\ &= (\sigma^{k-1}(\alpha_i^{r^{k-1}}))^s = \cdots = \alpha_i^{s^k}. \end{aligned}$$

Since α_i is non-zero it is a root of unity.

For the second claim, we put

$$f_p(x) := \prod_{i=1}^{d} (x - \alpha_i^p).$$

Then $f(x) := f_p(x) + pg(x)$ with f_p and $g \in \mathbb{Z}[x]$ by Fermat's little theorem. Further,

$$\begin{aligned} \left| \prod_{i=1}^{d} f(\alpha_i^p) \right| &= \left| \prod_{i=1}^{d} (f_p(\alpha_i^p) + pg(\alpha_i^p) \right| \\ &= p^d \left| \prod_{i=1}^{d} g(\alpha_i^p) \right|. \end{aligned}$$

Since α is not a root of unity by the first part of the lemma,

$$\prod_{i=1}^{d} f(\alpha_i^p) \neq 0$$

and so

$$\prod_{i=1}^{d} g(\alpha_i)$$

is non-zero. But $\prod_{i=1}^{d} g(\alpha_i)$ is an integer and so at least one in absolute value and the result follows. □

Theorem 21 (Dobrowolski)**.** *There exists a positive number c such that if α is an algebraic number of degree $d > 3$ and α is not a root of unity then*

$$\mathcal{M}(\alpha) > 1 + c \left(\frac{\log \log d}{\log d} \right)^3.$$

Proof. We first construct by means of Siegel's lemma and the assumption that

$$\mathcal{M}(\alpha) < 1 + \frac{1}{200}\left(\frac{\log\log d}{\log d}\right)^3. \tag{29}$$

Let F be a polynomial with small integer coefficients which is divided by a high power of f. Put

$$N = d\left\lfloor\frac{4\log d}{\log\log d}\right\rfloor^2$$

and

$$M = \left\lfloor\frac{4\log d}{\log\log d}\right\rfloor,$$

and suppose that d is large enough so that

$$\frac{\log d}{\log\log d} > 1.$$

The result is immediate if α is not an algebraic integer so we may suppose that α is an algebraic integer. Suppose that $\alpha = \alpha_1, \ldots, \alpha_d$ are the conjugates of α and that

$$f(x) = \prod_{i=1}^{d}(x - \alpha_i)$$

is the minimal polynomial of α. Put

$$F(x) = \sum_{i=1}^{N} a_i x^i,$$

and consider the M equations

$$\begin{aligned}
F(\alpha) &= \sum_{i=1}^{N} a_i\alpha^i = 0 \\
F^1(\alpha) &= \sum_{i=1}^{N} i a_i\alpha^{i-1} = 0 \\
&\vdots \\
F^{(m-1)}(\alpha) &= \sum_{i=m-1}^{N} i(i-1)(i-2)\cdots(i-M+2)a_i\alpha^{i-M+1} = 0.
\end{aligned} \tag{$*$}$$

We apply Siegel's lemma II to find integers $a_1, \ldots, a_N$ not all zero for which $(*)$ holds satisfying

$$\begin{aligned}
\max_i |a_i| &\le \left(2\sqrt{2}(N+1)(N^{(1+2+\cdots+M)d}\mathcal{M}(\alpha)^{NM})^{\frac{1}{dM}}\right)^{\frac{d\left\lfloor\frac{4\log d}{\log\log d}\right\rfloor}{\frac{99}{100}d\left\lfloor\frac{4\log d}{\log\log d}\right\rfloor^2}} \\
&\le \left(2\sqrt{2}(N+1)^{\frac{M+1}{2}}\mathcal{M}(\alpha)^{\frac{N}{d}}\right)^{\frac{100/99}{\lfloor 4\log d/\log\log d\rfloor}},
\end{aligned}$$

and since $2\sqrt{2}(N+1) < N^{3/2}$ for d sufficiently large, we have

$$\begin{aligned}\max_i |a_i| &< \left(N^{\frac{M+4}{2}}\mathcal{M}(\alpha)^{\frac{N}{d}}\right)^{\frac{100}{99}\cdot\frac{1}{\left\lfloor\frac{4\log d}{\log\log d}\right\rfloor}}\\ &< N^{11/20}\mathcal{M}(\alpha)^{5\log d/\log\log d}.\end{aligned}$$

Recalling (29) we see that for d sufficiently large we have $\max_i |a_i| < N^{3/5}$. We let F be defined by the a_i so that $f(x)^M \,|\, F(x)$. We will show that F has many other zeroes. In fact, too many! Let p be a prime with

$$\left(\frac{\log d}{\log\log d}\right)^2 < p < \frac{40(\log d)^2}{\log\log d}.$$

Then $F(x) = f(x)^M g(x)$ with $g \in \mathbb{Z}[x]$. We claim that α^p is a root of F. To see this, note that

$$|\operatorname{N}_{\mathbb{Q}(\alpha)/\mathbb{Q}} F(\alpha^p)| = \left|\prod_{i=1}^{d} F(\alpha_i^p)\right| \leq (NY)^d \mathcal{M}(\alpha)^{Np}, \tag{30}$$

where $Y := N^{2/3}\mathcal{M}(\alpha)^{6\log d/\log\log d}$. On the other hand, by Lemma 3,

$$\left|\prod_{i=1}^{d} f(\alpha_i^p)\right| \geq p^d$$

provided that α is not a root of unity. Assume that α is not a root of unity. Then since $F(x) = f(x)^M g(x)$, either α^p is a root of F or

$$|\operatorname{N}_{\mathbb{Q}(\alpha)/\mathbb{Q}} F(\alpha^p) \geq p^{dM}. \tag{31}$$

Comparing (30) and (31) we find that

$$p^{dM} \leq (NY)^d \mathcal{M}(\alpha)^{Np},$$

so we have

$$p \leq N^{8/5}\mathcal{M}(\alpha)^{\frac{N}{d}p}.$$

But $p^M \geq e^{\log\log d\left(\frac{4\log d}{\log\log d}\right)} = d^4$ for d sufficiently large, whereas $N^{8/5} < d^2$ for d sufficiently large. Therefore

$$P^{M/2} \leq \mathcal{M}(\alpha)^{\frac{N}{d}p}.$$

But then

$$\frac{M}{2}\log p \leq \frac{N}{d} p \log \mathcal{M}(\alpha),$$

so

$$\begin{aligned}1 &\leq \frac{2N}{dM}\frac{p}{\log p}\log\mathcal{M}(\alpha)\\ &\leq \frac{8\log d}{\log\log d}\cdot 22\left(\frac{\log d}{\log\log d}\right)^2\log\mathcal{M}(\alpha)\end{aligned}$$

for d sufficiently large. Therefore

$$\frac{1}{176}\left(\frac{\log\log d}{\log d}\right)^3 \leq \log\mathcal{M}(\alpha). \tag{32}$$

But by (29) and the inequality $\log(1+x) < x$ for $x > 0$. We see that

$$\log \mathcal{M}(\alpha) < \frac{1}{200}\left(\frac{\log\log d}{\log d}\right)^3,$$

which contradicts (32). Therefore, since d is not a root of unity, α^p is a root of F. But then $\alpha_1^p, \dots, \alpha_d^p$ are roots of F. Further, by Lemma 3, for distinct primes p_1 and p_2 we have $\alpha_i^{p_1} \neq \alpha_j^{p_2}$ for all i and j. Further by Lemma 2, $\alpha_1^p, \dots, \alpha_d^p$ are distinct for all but $\frac{\log d}{\log 2}$ primes. The number of primes which contribute d distinct roots is at least

$$\pi\left(\frac{40(\log d)^2}{\log\log d}\right)^2 - \pi\left(\left(\frac{\log d}{\log\log d}\right)^2\right) - \frac{\log d}{\log 2}$$

which by the prime number theorem is at least

$$18\left(\frac{\log d}{\log\log d}\right)^2$$

for d sufficiently large. On the other hand, we have

$$N \leq 16d\left(\frac{\log d}{\log\log d}\right)^2$$

which is a contradiction. Thus the result holds for d sufficiently large and so for $d > 3$. □

18. November 2

Remark 8. Ideas of the proof goes as follows:

(1) Construct a polynomial $F \in \mathbb{Z}[x]$ with "small" coefficients divisible by a large power of the minimal polynomial of α.
(2) Show that F has zero at α^p for primes p "not too large".
(3) Show that these give many new zeroes of F.
(4) Then we get too many new zeroes if α is not a root of unity.
(5) Note that all steps under the assumption that $\mathcal{M}(\alpha)$ is "small".

Time to return to the approximation of algebraic numbers by rationals.

Theorem 22. *Let α be an algebraic number of degree $d > 1$. This is an effectively computable positive number $C(\alpha)$, which depends on α, such that*

$$\left|\alpha - \frac{p}{q}\right| > \frac{C(\alpha)}{q^d},$$

for every rational p/q with $q > 0$.

Proof. Let f be the minimal polynomial of α over the integers. We may assume that α is real since otherwise we can take

$$C(\alpha) = \min_{x \in \mathbb{R}} |\alpha - x|.$$

Then, since α is not rational, $f(p/q) \neq 0$. Thus by the mean value theorem, we have

$$\frac{1}{q^d} \leq \left|f\left(\frac{p}{q}\right)\right| = \left|f(\alpha) - f\left(\frac{p}{q}\right)\right| \leq \left|\alpha - \frac{p}{q}\right| |f'(\theta)| \tag{33}$$

where θ is a real number between $\frac{p}{q}$ and α. Note that the results holds if $|\alpha-\frac{p}{q}|>1$ so we may assume that $|\alpha-\frac{p}{q}|\leq 1$. Suppose $f(x)=a_dx^d+\cdots+a_1x+a_0$. Then $f'(x)=da_dx^{d-1}+\cdots+a_1$ and so

$$|f'(\theta)|\leq d|a_d|(|\alpha|+1)^{d-1}+\cdots+|a_1|$$

and by (33) the results holds with $C(\alpha^{-1})$ equal to $d|a_d|(|\alpha|+1)^{d-1}+\cdots+|a_1|$. □

As we mentioned previously, an immediate consequence is that

$$\gamma:=\sum_{n=1}^{\infty}10^{-n!}$$

is transcendental. To see this, take

$$p_k:=10^{k!}\sum_{n=1}^{k}10^{-n!}, q_k=10^{k!}.$$

Then

$$\left|\gamma-\frac{p_k}{q_k}\right|<\frac{2}{10^{(k+1)!}}<\frac{2}{q_k^{k+1}}. \qquad (*)$$

If γ is algebraic of degree d then

$$\left|\gamma-\frac{p_k}{q_k}\right|>\frac{C(\gamma)}{q_k^d}$$

for some $C(\theta)>0$. But then by $(*)$ we have $q_k^{k+1-d}<2C(\gamma)^{-1}$. This gives contradiction for sufficiently large k.

19. November 4

Let α be an algebraic number of degree $d>1$. Consider the inequality

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{q^\mu}. \qquad (34)$$

Liouville's result showed that (34) has only finitely many solutions $\frac{p}{q}$ with $\mu>d$. Thue in 1908/1909 showed that (34) has finitely many solutions $\frac{p}{q}$ with $\mu>\frac{d}{2}+1$. Siegel in 1921 improved it to $\mu(2\sqrt{d})$. Dyson showed the same result for $\mu>\sqrt{2d}$. Roth in 1955 showed that the result holds for $\mu>2$. However, none of the proven results are effective.

Theorem 23 (Roth's theorem). *Let α be an algebraic number and let δ be a positive real number. There are only finitely many distinct rationals $\frac{p}{q}$ with $q>0$ for which*

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{q^{2+\delta}}. \qquad (*)$$

Motivated by the typical behaviour of approximation to a real number one might conjecture – an indeed Lang has – that $(*)$ holds with $\frac{1}{q^{2+\delta}}$ replaced by $\frac{1}{q^2(\log q)^{1+\delta}}$. However there has been no improvement on $(*)$ obtained yet. Notice that Roth's theorem tells us something about the growth of the partial quotient of α. In particular, $a_{n+1}<q_n^\delta$, for all but finitely many n's. Note that $q_0=1, q_1=a_1q_0+1,\ldots,q_n=a_nq_{n-1}+q_{n-2}$. Thus for $n\geq 2$, we have $q_n<(a_1+1)\cdots(a_n+1)$, and so $a_n<((a_1+1)(a_2+1)\cdots(a_n+1))^{2\delta}$, for n sufficiently large. It follows from this observation that $\log\log q_n<C_1(\alpha)n$ for a positive number $C_1(\alpha)$.

In 1955, Davenport and Roth did slightly better. They proved that for all real algebraic irrationals we have

$$\log\log q_n < C_2(\alpha)\frac{n}{\sqrt{\log n}}.$$

Perhaps the most important consequence of Roth's theorem is its use in the study of Diophantine equations. Let m be a positive integer and let

$$F(x,y) = a_d x^d + a_{d-1}x^{d-1}y + \cdots + a_0 y^d$$

be a binary form of degree $d > 2$ with non-zero discriminant. The equation

$$F(x,y) = m$$

in integers x and y is known as a Thue equation. For example, $x^3 - 2y^3 = 6$ is a Thue equation. Over $\mathbb{C}$ we can factor F in the following manner: $F(x,y) = L_1(x,y)L_2(x,y)\ldots L_d(x,y)$ where $L_i(x,y) = \gamma_i x + \delta_i y$ for $i = 1,\ldots,d$. Factor with the γ_i and δ_i's algebraic numbers. Since F has non-zero discriminant, any two linear forms L_i and L_j with $i \neq j$ are linearly independent over $\mathbb{C}$. Let (x,y) be a solution of $F(x,y) = m$. We may order the linear forms so that

$$0 < |L_1(x,y)| \leq |L_2(x,y)| \leq \cdots \leq |L_d(x,y)|.$$

(Since $m \neq 0, |L_1(x,y:)| > 0$.) Now if $\gamma_1 = 0$ or δ_1/γ_1 is in $\mathbb{Q}$, then $|L_1(x,y)| \geq c_1$ for some positive constant c_1. If $\gamma_1 \neq 0$ and $y = 0$ then $|L_1(x,y)| = |\gamma_1|(|x| + |y|)$. Finally, if $\gamma_1 \neq 0$, δ_1/γ_1 is irrational and $y \neq 0$ then

$$L_1(x,y) = \gamma_1 y\left(\frac{x}{y} - \alpha\right)$$

where

$$\alpha = -\frac{\delta_1}{\gamma_1}.$$

For every $\varepsilon > 0$ we have, by Roth's theorem, that

$$|L_1(x,y)| \geq C_2(a,\varepsilon)\frac{1}{|y|^{1+\varepsilon}} \geq C_2(\alpha,\varepsilon)(|x| + |y|)^{-1-\varepsilon}.$$

Since L_1 and L_2 are linearly independent over $\mathbb{C}$, we have

$$|L_2(x,y)| \geq \frac{1}{2}(|L_1(x,y)| + |L_2(x,y)|) \geq C_3(|x| + |y|).$$

Thus

$$\begin{aligned}|F(x,y)| &\geq C_2(\alpha,\varepsilon)C_3^{d-1}(|x| + |y|)^{d-1}(|x| + |y|)^{-1-\varepsilon}\\ &\geq C_4(\alpha,\varepsilon)(|x| + |y|^{d-2-\varepsilon}.\end{aligned}$$

Therefore if F is a binary form with integer coefficients and with non-zero discriminant and with degree $d \geq 3$, then there are only finitely many integer pairs (x,y) with

$$0 < |F(x,y)| < (|x| + |y|)^{\theta}$$

where θ is a real number with $0 < \delta < d - 2$. In particular, the Thue equation has only finitely many solutions.

20. November 6

A vast generalization of Roth's theorem is Schmidt's subspace theorem. To se the stage for it we first state, without proof, a generalization of Liouville's theorem.

Theorem 24. *Suppose that $1, \alpha_1, \ldots, \alpha_n$ are real algebraic numbers which are linearly independent over $\mathbb{Q}$ and the degree $\mathbb{Q}(\alpha_1, \ldots, \alpha_n)$ over $\mathbb{Q}$ is d. Then there is a positive number c such that*

$$|\alpha_1 q_1 + \cdots + \alpha_n q_n - p| > c q^{-d+1},$$

for all integers $q_1, \ldots, q_n$ and p with $q := \max |q_i| > 0$. Notice that if $n = 1$ then we recover Liouville's theorem.

Theorem 25 (Schmidt)**.** *Let $1, \alpha_1, \ldots, \alpha_n$ be real algebraic numbers which are linearly independent over $\mathbb{Q}$. Let $\delta > 0$. Then there are only finitely many n-tuple of non-zero integers $q_1, \ldots, q_n$ with*

$$|q_1 q_2 \cdots q_n|^{1+\delta} \|\alpha_1 q_1 + \cdots + \alpha_n q_n\| < 1.$$

Apply Theorem 25 to all the non-empty subsets of $\{\alpha_1, \ldots, \alpha_n\}$ Schmidt obtained.

Corollary 1. *Let $1, \alpha_1, \ldots, \alpha_n$ be real algebraic numbers which are linearly independent over $\mathbb{Q}$. Let $\delta > 0$. There are only finitely many $(n+1)$-tuple of integers $q_1, \ldots, q_n$ and p with $q := \max |q_i| > 0$ for which*

$$|\alpha_1 q_1 + \cdots + \alpha_n q_n - p| < \frac{1}{q^{n+\delta}}.$$

Schmidt also proved the following result.

Theorem 26. *Suppose $\alpha_1, \ldots, \alpha_n$ are real algebraic numbers with $1, \alpha_1, \ldots, \alpha_n$ linearly independent over $\mathbb{Q}$. Let $\delta > 0$. Then there are only finitely many positive integers q with*

$$q^{1+\delta} \|\alpha_1 q\| \cdots \|\alpha_n q\| < 1.$$

As an immediate consequence of Theorem 26 we have

Corollary 2. *Let $\alpha_1, \ldots, \alpha_n$ be real algebraic numbers with $1, \alpha_1, \alpha_2, \ldots, \alpha_n$ linearly independent over $\mathbb{Q}$ and let $\delta > 0$. Then there are only finitely many rational n-tuples $\left(\frac{p_1}{q}, \ldots, \frac{p_n}{q}\right)$ with*

$$\left|\alpha_i - \frac{p_i}{q}\right| < \frac{1}{q^{1+n^{-1}+\delta}}$$

for $i = 1, \ldots, n$.

Definition 20. We define the *house of* $\mathbf{x}$ (written $\overline{|\mathbf{x}|}$) as

$$\overline{|\mathbf{x}|} := \max_i |x_i|.$$

We will deduce Theorems 25 and 26 from a result proved by Schmidt in 1972.

Theorem 27 (Schmidt subspace theorem)**.** *Suppose $L_1(\mathbf{x}), \ldots, L_n(\mathbf{x})$ are linearly independent linear forms in $\mathbf{x} := (x_1, \ldots, x_n)$ with (real or complex) algebraic coefficients. Let $\delta > 0$. There are finitely many proper subspaces $T_1, \ldots, T_w$ of $\mathbb{R}^n$ such that every integer point $\mathbf{x}$ with $\mathbf{x} \neq \mathbf{0}$ and*

$$|L_1(\mathbf{x}) L_2(\mathbf{x}) \cdots L_n(\mathbf{x})| < \overline{|\mathbf{x}|}^{-\delta}$$

lies in one of these subspaces.

Remark 9. A few remarks on the subspace theorem:

(1) The result is *not* effective and so one cannot determine the subspaces $T_1, \dots, T_w$ from the proof.
(2) The integer points in a subspace T span a rational linear subspace. That is a subspace defined by linear equations with rational coefficients. Thus $T_1, \dots, T_2$ may be taken to be rational subspaces.
(3) The proof is difficult as it is a substantial generalization of Roth's theorem.

21. November 9

Let us deduce Theorem 26 from the subspace theorem.

Proof of Theorem 26. Suppose q is a positive integer fo which

$$q^{1+\delta}\|\alpha_1 q\| \cdots \|\alpha_n q\| < 1.$$

Let p_i be an integer for which $\|\alpha_i q\| = |\alpha_i q - p_i|$ for $i = 1, \dots, n$. Then put

$$(\mathbf{x}) = (x_1, x_2, \dots, x_{n+1}) = (p_1, \dots, p_n, q).$$

Let $C_1, C_2, \dots$ be positive numbers which depend on δ and $\alpha_1, \dots, \alpha_n$. Plainly we may take C_1 so that

$$\overline{|\mathbf{x}|} < C_1 q.$$

We consider the linear forms

$$L_i(\mathbf{X}) = \alpha_i X_{n+1} - X_i \text{ for } 1 \le i \le n$$

and

$$L_{n+1}(\mathbf{X}) = X_{n+1}.$$

Then

$$|L_1(\mathbf{x}) L_2(\mathbf{x}) \cdots L_{n+1}(\mathbf{x})| = \|\alpha_1 q\| \|\alpha_2 q\| \cdots \|\alpha_n q\| q,$$

so

$$|L_1(\mathbf{x}) \cdots L_{n+1}(\mathbf{x})| < q^{-\delta} < \overline{|\mathbf{x}|}^{\delta/2}$$

for q sufficiently large, as we may assume.

Then, by the subspace theorem, $\mathbf{x}$ lies in one of finitely many subspaces $T_1, \dots, T_w$. A typical subspace T is defined by

$$C_1 X_1 + \cdots + C_{n+1} X_{n+1} = 0$$

for $C_1, \dots, C_{n+1} \in \ Q$, not all zero. Then for $\mathbf{x} \in T$, we have

$$\begin{aligned} |C_1(\alpha_1 q - p_1) + \cdots + C_n(\alpha_n q - p_n)| &= |(C_1\alpha_1 + \cdots + C_n\alpha_n)q - (C_1 p_1 + \cdots + C_n p_n)| \\ &= |(C_1\alpha_1 + \cdots + C_n\alpha_n + C_{n+1})q| > C_2 q, \end{aligned}$$

since $1, \alpha_1, \dots, \alpha_n$ are $\mathbb{Q}$-linearly independent. Thus we have

$$C_2 q \le |C_1| \|\alpha_1 q\| + \cdots + |C_n| \|\alpha_n q\| \le |C_1| + \cdots + |C_n|,$$

hence q is bounded as required. □

Next we shall deduce Theorem 25 from the subspace theorem.

Proof. We shall prove the result by induction on n. The result when $n = 1$ follows from Theorem 26. Assume that $q_1, \ldots, q_n$ satisfy the hypothesis of Theorem 25. Choose p to be an integer so that

$$\|\alpha_1 q_1 + \cdots + \alpha_n q_n\| = |\alpha_1 q_1 + \cdots + \alpha_n q_n - p|.$$

Write $\mathbf{x} = (x_1, \ldots, x_{n+1}) = (q_1, \ldots, q_n, p)$. Then there exists a positive number C_3 which depends on $\alpha_1, \ldots, \alpha_n$ and δ only such that

$$\lceil\mathbf{x}\rceil < C_3 q,$$

where $q = \max_i |q_i|$. Put

$$L_i(\mathbf{X}) = X_i \text{ for } i = 1, \ldots, n,$$

and

$$L_{n+1}(\mathbf{X}) = \alpha_1 X_1 + \cdots + \alpha_n X_n - X_{n+1}.$$

Then

$$|L_1(\mathbf{x}) L_2(\mathbf{x}) \cdots L_{n+1}(\mathbf{x})| = |q_1 \ldots q_n| \|\alpha_1 q_1 + \cdots + \alpha_n q_n\|$$

so

$$|L_1(\mathbf{x}) \cdots L_{n+1}(\mathbf{x})| < \frac{1}{|q_1 \ldots q_n|^{\delta}} < \frac{1}{\lceil\mathbf{x}\rceil^{\delta/2}},$$

for q sufficiently large. Then by the subspace theorem $\mathbf{x}$ lies in one of finitely many rational subspaces. Let T be such a subspace containing $\mathbf{x}$. Then T is defined by

$$C_1 X_1 + \cdots + C_{n+1} X_{n+1} = 0,$$

with $C_1, \ldots, C_{n+1}$ in $\mathbb{Q}$ and not all zero. Then either one of $C_1, \ldots, C_n$ is non-zero or $C_1, \ldots, C_n$ are all zero and $C_{n+1} \neq 0$. In the first case, we may assume without loss of generality that $C_n \neq 0$. Let us now consider the firs case. Then

$$C_n q_n = -C_1 q_1 - \cdots - C_{n-1} q_{n-1} - C_{n+1} p,$$

so

$$C_n \alpha_n q_n = -C_1 \alpha_n q_1 - C_{n-1} \alpha_n q_{n-1} - C_{n+1} \alpha_n p.$$

Thus

$$\begin{aligned} |C_n||\alpha_1 q_1 + \cdots + \alpha_n q_n - p| &= |(C_n \alpha_1 - C_1 \alpha_n) q_1 + \cdots + (C_n \alpha_{n-1} C_{n-1} \alpha_n) q_{n-1} \\ &\qquad - (C_n + C_{n+1} \alpha_n) p| \\ &= |C_n + C_{n+1} \alpha_n| \left| \left(\frac{C_n \alpha_1 - C_1 \alpha_n}{C_n + C_{n+1} \alpha_n} \right) q_1 + \cdots + \right. \\ &\qquad \left. \left(\frac{C_n \alpha_{n-1} - C_{n-1} \alpha_n}{C_n + C_{n+1} \alpha_n} \right) q_{n-1} - p \right| \\ &= |C_n C_{n+1} \alpha_n| |\alpha'_1 q_1 + \cdots + \alpha'_{n-1} q_{n-1} - p|. \end{aligned}$$

Threfore, there exists a positive number C_4 which depends on $\alpha_1, \ldots, \alpha_n$ and δ such that

$$\|\alpha'_1 q_1 + \cdots + \alpha'_{n-1} q_{n-1}\| < \frac{C_4}{|q_1 \ldots q_n|^{1+\delta}} < \frac{1}{|q_1 q_2 \ldots q_{n-1}|^{1+\delta/2}},$$

for $q = \max_i |q_i|$ sufficiently large. To complete our induction we must check that $1, \alpha'_1, \ldots, \alpha'_{n-1}$ are $\mathbb{Q}$-linearly independent.

WE now check that $1, \alpha'_1, \ldots, \alpha'_{n-1}$ are $\mathbb{Q}$-linear independent. Observe that if $\lambda_1\alpha'_1 + \cdots + \lambda_{n-1} + \alpha'_{n-1} + \lambda_n = 0$ with $\lambda \in \mathbb{Q}$ for $i = 1, \ldots, n$. Then

$$\lambda_1(C_n\alpha_1 - C_1\alpha_n) + \cdots + \lambda_{n-1}(C_n\alpha_{n-1} - C_{n-1}\alpha_n) + \lambda_n(C_n + C_{n+1}\alpha_n) = 0.$$

Hence

$$\lambda_1 C_n\alpha_1 + \cdots + \lambda_{n-1}C_n\alpha_{n-1} - (\lambda_1 C_1 + \cdots + \lambda_{n-1}C_{n-1} + C_{n+1})\alpha_n = 0.$$

But then $\lambda_1 = \cdots = \lambda_n = -$ and so $1, \alpha'_1, \alpha'_2, \ldots, \alpha_{n-1}/$ are linearly independent over $\mathbb{Q}$. Then in this case, by induction $|q_1|, \ldots, |q_n|$ are bounded. It remains to consider the possibility that $C_1 = \cdots = C_n = 0$ and $C_{n+1} \neq 0$. Then $C_{n+1}p = 0$ hence $p = 0$ and in this case

$$|q_1 q_2 \cdots q_n|^{1+\delta}|\alpha_1 q_1 + \cdots + \alpha_n q_n| < 1.$$

In this case

$$|q_1 \cdots q_{n-1}|^{1+\delta}|\alpha_n| \left| \left(\frac{\alpha_1}{\alpha_n}\right) q_1 + \cdots + \left(\frac{\alpha_{n-1}}{\alpha_n}\right) q_{n-1} + q_n \right| < 1.$$

Put $\alpha'_i = \alpha_i/\alpha_n$ for $i = 1, \ldots, n-1$. Then

$$|q_1 q_2 \cdots q_{n-1}|^{1+\delta/2}|\alpha'_1 q_1 + \cdots + \alpha'_{n-1}q_{n-1} + q_n| < 1$$

for $q = \max_i |q_i|$ sufficiently large and the result again follows by induction. □

In a similar way we can deduce the following result from the subspace theorem:

Theorem 28. *Let α_{ij} be real algebraic numbers for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Suppose that $1, \alpha_{i1}, \ldots, \alpha_{im}$ are $\mathbb{Q}$-linearly independent for $i = 1, \ldots, n$. Let $\delta > 0$. Then there are only finitely many m-tuples of non-zero integers $(q_1, \ldots, q_m)$ for which*

$$|q_1 \cdots q_m|^{1+\delta} \prod_{i=1}^{n} \|\alpha_{i1}q_1 + \cdots + \alpha_{im}q_m\| < 1.$$

Instead of approximating algebraic numbers by rationals we can approximate by algebraic numbers.

Theorem 29. *Let n be a positive integer and $\varepsilon > 0$. If α is an algebraic number of degree greater than n then there are only finitely many algebraic numbers |beta of degree at most n for which*

$$|\alpha - \beta| < H_0(\beta)^{-n-1-\varepsilon}.$$

Recall that $H_0(\beta)$ denotes the naïve height of β.

Proof. We take $\alpha_j = \alpha^j$ for $j = 1, \ldots, m$ where m is the degree of β. Then $1, \alpha_1, \ldots, \alpha_j$ are linearly independent over $\mathbb{Q}$ since $m \leq n$. Let $P(x) = a_m x^m + \cdots + a_0$ be the minimal polynomial of β. We first note that if

$$P(x) = a_m(x - \beta_1) \cdots (x - \beta_m)$$

then

$$|P(\alpha)| = |a_m||\alpha - \beta_1| \cdots |\alpha - \beta_m|$$

where $\beta = \beta_1$. Thus

$$\begin{aligned}|P(\alpha)| &\leq |\alpha - \beta||a_m| \prod_{i=2}^{m} (2\max(|\alpha|, |\beta_i|)) \\ &\leq |\alpha - beta||a_m| \prod_{i=2}^{m} 2(\max(1, |\alpha|))(\max(1, |\beta_i|)) \\ &\leq |\alpha - \beta||a_m|(2\max(1, |\alpha|))^{m-1} \prod_{i=2}^{m} \max(1, |\beta_i|) \\ &\leq |\alpha - \beta| C_1 H_0(\beta), \qquad (35)\end{aligned}$$

where C_1 is a positive number which depends on α and n.

On the other hand, by the corollary to Theorem 25, for each $\varepsilon > 0$,

$$|P(\alpha)| > \frac{C_2(\alpha, n; \varepsilon)}{H_0(\beta)^{-n-\varepsilon}}. \qquad (36)$$

The result follows from (35) and (36). □

This result can be contrasted with Leveque's theorem.

Theorem 30 (Leveque)**.** *Let K be a finite extension over $\mathbb{Q}$ with $[K : \mathbb{Q}] = n$ and let α be algebraic of degree d over K. Let $\varepsilon > 0$. There are only finitely many $\beta \in K$ for which*

$$|\alpha - \beta| < H_0(\beta)^{-2-\varepsilon}.$$

Suppose that F is an irreducible binary form of degree d over $\mathbb{Q}$. Suppose the leading coefficient of $F(x, 1)$ is 1. Then

$$F(x, y) = (X - \alpha_1 Y) \cdots (X - \alpha_d Y),$$

and put $K = \mathbb{Q}(\alpha_1)$. Then $F(X, Y) = \mathrm{N}_{K/\mathbb{Q}}(X - \alpha_1 Y)$.

22. November 13

Let K be an algebraic number field of degree d over $\mathbb{Q}$. There are d isomorphic embeddings $\varphi_1, \ldots, \varphi_d$ of K into $\mathbb{C}$ which fix $\mathbb{Q}$. We denote the image of α in K under φ_i by $\alpha^{(i)}$ for $i = 1, \ldots, d$. Thus for $\alpha \in K$, we have

$$\mathrm{N}_{K/\mathbb{Q}}(\alpha) = \alpha^{(1)} \cdots \alpha^{(d)}.$$

Given a linear form

$$M(\mathbf{x}) = \alpha_1 X_1 + \cdots + \alpha_n X_n,$$

with $\alpha_i \in K$ for $i = 1, \ldots, n$. We write

$$\mathrm{N}(M(\mathbf{x})) = \prod_{i=1}^{d} M^{(i)}(\mathbf{x}) = \prod_{i=1}^{d} (\alpha_1^{(i)} X_1 + \cdots + \alpha_n^{(i)} X_n).$$

Definition 21. A form $F(\mathbf{x})$ with $F(X) = N(M(\mathbf{x}))$ for some linear form M with coefficients in K is said to be a *norm form.*

Thus for example if $\sqrt[4]{2}$ and $M(X) = X_1 - \sqrt[4]{2}X_2$ then $N(M(\mathbf{x})) = X_1^4 - 2X_2^4$, whereas if

$$M(\mathbf{x}) = X_1 + \sqrt[4]{2}X_2 + \sqrt[4]{4}X_3,$$

then

$$N(M(\mathbf{x})) = X_1^4 - 2X_2^4 + 4X_3^4 - 4X_1^2X_3^2 + 8X_1X_2^2X_3.$$

Note that a norm form is a homogeneous polynomial of degree $[K : \mathbb{Q}]$. If the coefficients $\alpha_1, \dots, \alpha_n$ of the linear form $M(\mathbf{x})$ are algebraic integers then the coefficients of the norm forms are integers. Let a be a non-zero integer. The equation

$$N(M(\mathbf{x})) = a \tag{37}$$

in integers $x_1, \dots, x_n$ is known as a norm form equation. Put $\mathcal{M} = \{M(\mathbf{x}) : \mathbf{x} \in \mathbb{Z}^n\}$. $\mathcal{M}$ is a $\mathbb{Z}$-module since $\mathcal{M}$ is an additive abelian group; and for all $r \in \mathbb{Z}$ and $m \in \mathcal{M}$ we have $rm \in \mathcal{M}$. And for all $r, s \in \mathbb{Z}$ and $m, n \in \mathcal{M}$,

- $r(m + n) = rm + rn$
- $(r + s)m = rm + sm$
- $r(sm) = (rs)m.$
- $1 \cdot m = m.$

Thus we can view the norm form equation (37) equivalently as $\mathrm{N}(\mu) = a$ with $\mu \in \mathcal{M}$. We will now show that a module $\mathcal{M}$ in K has a basis. That is a system of generators which is $\mathbb{Z}$-linearly independent, i.e., $a_1\alpha_1 + \cdots + a_n\alpha_n = 0$ with $a_i \in \mathbb{Z}$ for $i = 1, \dots, n$ then $a_1 = \cdots = a_n = 0$. We will deduce this from the following result on abelian groups.

Theorem 31. *If an abelian group has no non-zero element of finite order and it possesses a finite system of generators then it possesses a basis.*

Proof. Let $\alpha_1, \dots, \alpha_s$ be a system of generators of the group M so $M = (\alpha_1, \dots, \alpha_s)$. Note that for any $k \in \mathbb{Z}$ we have

$$M = (\alpha_1 + k\alpha_2, \dots, \alpha_2, \dots, \alpha_s),$$

since if $\alpha \in M$ and $\alpha_1' = \alpha_1 + k\alpha_2$ then

$$\alpha = c_1\alpha_1 + \cdots + c_s\alpha_s = c_1\alpha_1' + (c_2 - kc_1)\alpha_2 + \cdots + c_s\alpha_s,$$

so α is an integer linear combination of $\alpha_1', \dots, \alpha_s$. If $\alpha_1, \dots, \alpha_s$ are $\mathbb{Z}$-linearly independent then they form a basis. If they are linearly dependent then

$$c_1\alpha_1 + \cdots + c_s\alpha_s = 0 \tag{38}$$

with the c_i's integers and not all zero. Suppose, without loss of generality, that $c_1 \neq 0$ and that c_1 has the smallest non-zero absolute value. Suppose that c_1 does not divide all the other c_i's. Without loss of generality, we may suppose that $c_1 \nmid c_2$. Then $c_2 = qc_1 + r$ with $0 < r < |c_1|$. We now consider the system of generators $\alpha_1' = \alpha_1 + q\alpha_2, \alpha_2, \dots, \alpha_s$. Then (38) becomes

$$c_1\alpha_1' + r\alpha_2 + c_3\alpha_3 + \cdots + c_s\alpha_s = 0. \tag{39}$$

Therefore the generators are linked by a relation with a coefficient which is non-zero and smaller in absolute value then $|c_1|$. We now repeat the argument. After at most $|c_1|$ steps we must arrive at a system of generators $\beta_1, \dots, \beta_s$ and a relation

$$k_1\beta_1 + \cdots + k_s\beta_s = 0$$

where $k_1, \dots, k_s$ are not all zero, $k_1 \neq 0$ and $k_1 \,|\, k_i$ for all $i = 1, \dots, s$. Thus $\beta_1 + l_2\beta_2 + \cdots + l_s\beta_s = 0$, where

$$l_i = \frac{k_i}{k_1}$$

for $i = 1, \dots, s$ since 0 is the only element of finite order in the group. Thus we may express β_1 as an integer linear combination of $\beta_2, \dots, \beta_s$ so $M = (\beta_2, \dots, \beta_s)$. We now repeat the argument with $\beta_2, \dots, \beta_s$ if $\beta_2, \dots, \beta_s$ are not $\mathbb{Z}$-linearly independent. Eventually, after a finite number of steps, we eventually arrive at a basis. □

23. November 18

Some remarks on the connection between the height and the naive height of an algebraic number α: Recall that α is of degree d with conjugates $\alpha_1 = \alpha, \dots, \alpha_d$ and minimal polynomial f over the integers, say

$$f(x) = a_d x^d + \cdots + a_1 x + a_0,$$

then the naive height $H_0(\alpha)$ is given by $H_0(\alpha) = \max(|a_d|, \dots, |a_0|)$ and the height $H(\alpha)$ is given by the positive real number $H(\alpha)$ satisfying

$$H(\alpha)^d = \mathcal{M}(f) = |a_d| \prod_{i=1}^{d} \max(1, |\alpha_i|).$$

Recall from Jensen's theorem that

$$\begin{aligned} H(\alpha)^d &= \exp\left(\int_0^1 \log|f(e^{2\pi i\theta})|\, d\theta\right) \\ &\leq \exp\left(\int_0^1 \log(|a_d| + \cdots + |a_0|t)\, d\theta\right) \\ &\leq \exp\log(|a_d| + |a_{d-1}| + \cdots + |a_0|) \\ &\leq |a_d| + \cdots + |a_0| \\ &\leq (d+1)H_0(\alpha). \end{aligned}$$

Further, we have

$$H_0(\alpha) \leq (2H(\alpha))^d,$$

on noting that the a_j's are elementary symmetric functions in the conjugates of α. The j-th such function has $\binom{d}{j}$ terms, each smaller in absolute value than $H(\alpha)^d$. Further we have $\binom{n}{j} \leq 2^d$.

In 1900, Hilbert produced a list of 23 problems which he felt were of fundamental importance for mathematics for the international congress in Paris. His tenth problem was the following: can one find a universal method or algorithm for determining if a polynomial equation $f(x_1, \dots, x_n) = 0$ with $f \in \mathbb{Z}[x_1, \dots, x_n]$ has a solution in integers $x_1, \dots, x_n$?

The answer is *no*, and this was proved by Matiyasevich in 1970 building on work of Davis, Putnam, and Robinson.The same result has been obtained for certain rings of algebraic integers in place of $\mathbb{Z}$. However, the answer is not known if we replace $\mathbb{Z}$ with $\mathbb{Q}$. One consequence of Matiyasevich's work was that one could produce polynomials which had the property that every prime occurred exactly one time as a positive value of the polynomial and no other integers were positive values. Here is an example of such polynomial:

$$\begin{aligned} F(a,b,\dots,z) = {} & (k+2)(1-(wz+h+j-g)^2-(2n+p+q+z-e)^2-(a^2y^2-y^2+1-x^2)^2 \\ & -(\{e^4+2e^3\}(a+1)^2+1-o^2)^2-(16(k+1)^3(k+2)(n+1)^2+1-f^2)^2 \\ & -(\{(a+u^4-u^2a)^2-1\}\{n+4dy\}^2+1-\{x+cu\}^2)^2-(ai+k+1-l-i)^2 \\ & -(\{gk+g+k+1\}\{h+j\}+h-z)^2-(16r^2y^4\{a^2-1\}+1-u^2)^2 \\ & -(p-m+l\{a-n-1\}+b\{2an+2a-n^2-2n-2\})^2 \\ & -(z-pm+pla-p^2l+t\{2ap-p^2-1\})^2 \\ & -(q-x+y(a-p-1)+s(2ap+2a-p^2-2p-2)^2 \\ & -(a^2l^2-l^2+1-m^2)^2-(n+l+v-y)^2). \end{aligned}$$

One striking consequence is that one can give a certificate for a number of to be prime, i.e., values for $a, b, \dots, z$ which can be verified with only a fixed constant number of additions and multiplications. Of course, the values $a, \dots, z$ may be large in terms of the size of the prime.

24. November 20

Let K be a finite extension of $\mathbb{Q}$. Since K has characteristic 0 there are no non-trivial divisors of zero under addition. By Theorem 31 if $\mathcal{M}$ is a finitely-generated $\mathbb{Z}$-module in K then $\mathcal{M}$ has a basis. Further the maximum number of linearly independent terms of $\mathcal{M}$ over $\mathbb{Q}$ is $[K:\mathbb{Q}]$. Therefore every basis for a finitely-generated $\mathbb{Z}$-module in K has at most $[K:\mathbb{Q}]$ basis elements. The number of generators in a basis for such a module $\mathcal{M}$ is said to be the rank of $\mathcal{M}$. This is well-defined since any two bases for $\mathcal{M}$ have the same number of elements. Notice that if $\alpha_1, \dots, \alpha_m$ and $\alpha_1', \dots, \alpha_m'$ are bases for $\mathcal{M}$ then one basis can be transformed to another by an $m \times m$ unimodular matrix. In particular, by a matrix with integer entries and determinant ± 1.

Definition 22. We say that a module $\mathcal{M}$ in K is *full* if its rank is equal to $[K:\mathbb{Q}]$.

Theorem 32. *The norm form* $\mathrm{N}(\alpha_1 X_1 + \cdots + \alpha_n X_n)$ *is irreducible over the rationals if and only if* $K = \mathbb{Q}\left(\frac{\alpha_2}{\alpha_1}, \dots, \frac{\alpha_n}{\alpha_1}\right)$.

Proof. ($\Rightarrow$) We have $K = \mathbb{Q}(\alpha_1, \dots, \alpha_n)$ and $\mathrm{N}(\alpha_1 X_1 + \cdots + \alpha_n X_n) = \mathrm{N}_{K/\mathbb{Q}}(\alpha_1 X_1 + \cdots + \alpha_n X_n)$. Further,

$$\mathrm{N}(\alpha_1 X_1 + \cdots + \alpha_n X_n) = \mathrm{N}(\alpha_1)\,\mathrm{N}\left(X_1 + \frac{\alpha_2}{\alpha_1}X_2 + \cdots + \frac{\alpha_n}{\alpha_1}X_n\right).$$

Put $L := \mathbb{Q}\left(\frac{\alpha_2}{\alpha_1}, \dots, \frac{\alpha_n}{\alpha_1}\right)$. Then we have

$$\begin{aligned} \mathrm{N}(\alpha_1 X_1 + \cdots + \alpha_n X_n) &= \mathrm{N}(\alpha_1)\,\mathrm{N}_{K/\mathbb{Q}}\left(X_1 + \frac{\alpha_2}{\alpha_1}X_2 + \cdots + \frac{\alpha_n}{\alpha_1}X_n\right) \\ &= \mathrm{N}(\alpha_1)\,\mathrm{N}_{L/\mathbb{Q}}\left(X_1 + \frac{\alpha_2}{\alpha_1}X_2 + \cdots + \frac{\alpha_n}{\alpha_1}X_n\right)^{[K:L]}. \end{aligned}$$

Thus so if $\mathrm{N}(\alpha_1 X_1 + \cdots + \alpha_n X_n)$ is irreducible over $\mathbb{Q}$ then $[K : L] = 1$ so $K = L$.

($\Leftarrow$) On the other hand, if $K = L$ then by the primitive element theorem we have $K = \mathbb{Q}(\beta)$ for some β with

$$\beta = \alpha_1^{-1} \sum_{i=2}^{n} c_i \alpha_i,$$

where $c_2, \ldots, c_n \in \mathbb{Q}$. Let $[K : \mathbb{Q}] = d$. Then the degree of β over $\mathbb{Q}$ is d and so the binary form $\mathrm{N}(X + \beta Y)$ is irreducible over $\mathbb{Q}$. Thus

$$\begin{aligned} \mathrm{N}(X + \beta Y) &= \mathrm{N}\left(X + c_2 \frac{\alpha_2}{\alpha_1} Y + \cdots + c_n \frac{\alpha_n}{\alpha_1} Y\right) \\ &= \mathrm{N}\left(X_1 + \frac{\alpha_2}{\alpha_1} X_2 + \cdots + \frac{\alpha_n}{\alpha_1} X_n\right) \\ &= \mathrm{N}(\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n) \end{aligned}$$

is irreducible over $\mathbb{Q}$. □

Definition 23. Let $[K : \mathbb{Q}] < \infty$. A full $\mathbb{Z}$-module $\mathcal{M}$ of K which contains 1and is a ring is called an *order of* K.

The ring of algebraic integers of K is an order of K. Notice that if $\mathcal{O}$ is an order of K and $\mu \in \mathcal{O}$ then $\mu^h \in \mathcal{O}$ for h a positive integer. For each $\mathbb{Z}$-module $\mathcal{M}$ of K we can find a non-zero integer c such that cm is an algebraic integer for every $m \in \mathcal{M}$. Therefore take such a c for $\mathcal{O}$ and observe that $c\mu^h$ is an algebraic integer for $h = 1, 2, \ldots$. Therefore μ is an algebraic integer. Thus every order $\mathcal{O}$ of K is contained in the ring of algebraic integers of K. For this reason we call the ring of algebraic integers of K the *maximal order of* K.

The units in an order $\mathcal{O}$ are the divisors of 1. Note that if ε is a unit in $\mathcal{O}$ then $\varepsilon\varepsilon_1 = 1$ with $\varepsilon_1 \in \mathcal{O}$. Further $1 = \mathrm{N}(\varepsilon\varepsilon_1) = \mathrm{N}(\varepsilon)\,\mathrm{N}(\varepsilon_1)$ and since ε and ε_1 are algebraic integers with $\mathrm{N}(\varepsilon) = \pm 1$. Further if $\mathrm{N}(\varepsilon) = \pm 1$ with $\varepsilon \in \mathcal{O}$ then ε is a root of its minimal polynomial over $\mathbb{Z}$, say $x^d + a_{d-1}x^{d-1} + \cdots + a_0 \in \mathbb{Z}[x]$. Then $\mathrm{N}(\varepsilon) = \pm a_0$. Therefore

$$\varepsilon^{d-1} + a_{d-1}\varepsilon^{d-2} + \cdots + a_1 = \pm \frac{\mathrm{N}(\varepsilon)}{\varepsilon},$$

which is in $\mathcal{O}$. Thus ε^{-1} is in $\mathcal{O}$ and hence ε is a unit in $\mathcal{O}$. Thus the units in $\mathcal{O}$ are the elements $\varepsilon \in \mathcal{O}$ with $\mathrm{N}(\varepsilon) = \pm 1$. The units in $\mathcal{O}$ form a group. In fact:

Proposition 6. *Let $\mathcal{O}$ be an order in a finite extension K of $\mathbb{Q}$. The group of units it infinite except when $K = \mathbb{Q}$ or K is an imaginary quadratic extension of $\mathbb{Q}$.*

Proof. This is an extension of Dirichlet's unit theorem to orders – see, for instance, Bouvich and Shafarevich. □

25. November 23

Proposition 7. *Let M be a finitely-generated abelian group with no non-zero element of finite order. All subgroups N of M have a finite number of generators and so possess a* *basis. Further, if $w_1, \ldots, w_n$ is a basis for M then there is a basis $\nu_1, \ldots, \nu_k$ of N of the form*

$$\begin{aligned}
\nu_1 &= c_{11}w_1 + \cdots + c_{1n}w_n \\
\nu_2 &= c_{22}w_2 + \cdots + c_{2n}w_n \\
&\vdots \\
\nu_k &= c_{kk}w_k + \cdots + c_{kn}w_n
\end{aligned}$$

where the c_{ij}'s are integers and $c_{ii} > 0$ for $i = 1, \ldots, k$ and $k \leq m$.

Proof. Standard argument. □

Thus submodule of a module of K is a module of K, i.e., is a finitely-generated $\mathbb{Z}$-module. Given a full module $\mathcal{M}$ of K, let $\mathcal{O}_{\mathcal{M}}$ denote the set of $\lambda \in K$ such that $\lambda\mathcal{M} \subseteq \mathcal{M}$, i.e., $\lambda\mu \in \mathcal{M}$ for all $\mu \in \mathcal{M}$. $\mathcal{O}_M$ is known as the stabilizer of $\mathcal{M}$ or the coefficient ring of $\mathcal{M}$.

Proposition 8. *Let $[K : \mathbb{Q}] < \infty$. If $\mathcal{M}$ is a full module of K then $\mathcal{O}_{\mathcal{M}}$ is an order of K.*

Proof. The set $\mathcal{O}_{\mathcal{M}}$ is a ring since it is a non-empty subset of K, contains 1 and if $\theta_1, \theta_2 \in \mathcal{O}_{\mathcal{M}}$ then $\theta_1 + \theta_2$ and $\theta_1\theta_2$ are also in $\mathcal{O}_{\mathcal{M}}$. Next observe that $\mathcal{O}_M$ is a $\mathbb{Z}$-module since it is an abelian group under addition. In particular, for all $r \in \mathbb{Z}$ and for all $\theta \in \mathcal{O}_M$ we have $r\theta \in \mathcal{O}_{\mathcal{M}}$ since $\theta\mu\mathcal{O}_{\mathcal{M}}$ for all $\mu \in \mathcal{M}$, it follows $r\theta\mu \in \mathcal{O}_{\mathcal{M}}$. Further parts (i), (ii), (iii), and (iv) of the definition of a module hold. Thus $\mathcal{O}_{\mathcal{M}}$ is a module of K and is a ring with 1.

To prove that $\mathcal{O}_{\mathcal{M}}$ is an order we must show that $\mathcal{O}_{\mathcal{M}}$ is a full module in K. Let $\gamma \in \mathcal{M}$ with $\gamma \neq 0$. Then for all $\alpha \in \mathcal{O}_{\mathcal{M}}$ we have $\alpha\gamma \in \mathcal{M}$ hence $\gamma\mathcal{O}_{\mathcal{M}} \in \mathcal{M}$. Thus $\gamma\mathcal{O}_{\mathcal{M}}$ is a subgroup of $\mathcal{M}$ which is a module and so by Theorem 31 it possesses a basis and is finitely generated. Thus $\mathcal{O}_{\mathcal{M}} = \gamma^{-1}(\gamma\mathcal{O}_M)$ is finitely generated. Let $[K : \mathbb{Q}] = d$. To show that $\mathcal{O}_{\mathcal{M}}$ is full, it suffices to find d $\mathbb{Q}$-linearly independent in $\mathcal{O}_{\mathcal{M}}$. Start with $\alpha_1, \ldots, \alpha_d$ a basis for K over $\mathbb{Q}$. Let $\mathcal{M} = (\mu_1, \ldots, \mu_d)$, and recall that $\mathcal{M}$ is a full module. To test whether α in K is in $\mathcal{O}_{\mathcal{M}}$ it suffices to prove that $\alpha\mu_i$ is in $\mathcal{M}$ for $i = 1, \ldots, d$. Now we can write

$$\alpha\mu_i = \sum_{j=1}^{d}\sum_{j=1}^{d} a_{ij}\mu_j$$

with $a_{ij} \in \mathbb{Q}$ since $\mathcal{M}$ is full. For each α we can take c to be the least common multiple of the denominators of the a_{ij}'s. Then $ca_{ij} \in \mathbb{Z}$ for $1 \leq i \leq d, 1 \leq j \leq d$ and so $c\alpha \in \mathcal{O}_{\mathcal{M}}$. Thus for each integer i with $1 \leq i \leq d$, there is a non-zero integer c_i such that $c_i\alpha_i$ is in $\mathcal{O}_{\mathcal{M}}$. But then there are d $\mathbb{Q}$-linearly independent terms in $\mathcal{O}_{\mathcal{M}}$ so $\mathcal{O}_{\mathcal{M}}$ is full and the result follows. □

Let $[K : \mathbb{Q}] < \infty$. Let $\mathcal{M}$ be a full module of K. Let $U_{\mathcal{M}}$ denote the group of units in $\mathcal{O}_{\mathcal{M}}$ with norm 1. $U_{\mathcal{M}}$ is a subgroup of the group of units of $\mathcal{O}_{\mathcal{M}}$ of index 1 or 2, since the norm of a unit is ±1. Thus $U_{\mathcal{M}}$ is infinite except when K is $\mathbb{Q}$ or an imaginary quadratic extension of $\mathbb{Q}$ by Proposition 6 and Proposition 8. Now notice that if a is a non-zero integer and $\mu \in \mathcal{M}$ is a solution to

$$\mathrm{N}(\mu) = a, \tag{40}$$

then for all $\varepsilon \in U_{\mathcal{M}}$ where $\varepsilon\mu \in \mathcal{M}$ and

$$\mathrm{N}(\varepsilon\mu) = \mathrm{N}(\varepsilon)\,\mathrm{N}(\mu) = \mathrm{N}(\mu) = a.$$

Thus if $\mathcal{M}$ is a full module of K is not exceptional (so not $\mathbb{Q}$ or an imaginary quadratic extension of $\mathbb{Q}$) then (40) has infinitely many solutions whenever it has one non-zero solution.

26. November 25: Last lecture

Recall the norm form equation

$$\mathrm{N}(\mu) = a, \tag{41}$$

with $\mu \in \mathcal{M}$. Full modules in K are not the instances where we have infinitely many solutions of (41). Suppose that L is a subfield of K and $\mathcal{M}_0$ is a submodule of $\mathcal{M}$ which is proportional to a full module in L. In other words, $\mathcal{M}_0 = \gamma\mathcal{L}$ where $\mathcal{L}$ is a module in L and $\gamma \in K$ with γ non-zero. Now unless L is exceptional there will be infinitely many $\lambda \in L$ which satisfy

$$\mathrm{N}_{L/\mathbb{Q}}(\lambda) = b$$

for some non-zero integer b. But then $\mathrm{N}(\gamma\lambda) = \mathrm{N}(\gamma)\,\mathrm{N}(\lambda) = \mathrm{N}(\gamma)(\mathrm{N}_{L/\mathbb{Q}}(\lambda))^{[K:L]} = \mathrm{N}(\gamma)b^{[K:L]} = a$ for some $a \in \mathbb{Q}, a \neq 0$. Thus there are infinitely many solutions to (41) for some a in this situation also.

Definition 24. Let $[K : \mathbb{Q}] < \infty$. A module $\mathcal{M}$ of K is said to be *degenerate* if it contains a submodule which is proportional to a full module in some subfield L of K which is not $\mathbb{Q}$ or an imaginary quadratic extension of $\mathbb{Q}$.

We have shown that if $\mathcal{M}$ is degenerate then for certain values of a we have infinitely many solutions to (41).

Theorem 33 (Schmidt norm form theorem)**.** *Let K be a finite extension of $\mathbb{Q}$. Let $\mathcal{M}$ be a module of K. Then the following are equivalent:*

(1) there exists a non-zero a in $\mathbb{Q}$ for which the equation

$$\mathrm{N}(\mu) = a$$

has infinitely many solutions in μ in $\mathcal{M}$.

(2) $\mathcal{M}$ is degenerate.

Proof. ($\Leftarrow$) This one is straightforward.

($\Rightarrow$) Now *this* is the hard part – so hard that this proof is the beyond the scope of this lecture. However we shall remark that this direction follows from the Schmidt subspace theorem. The full proof will not be provided, however. □

Schmidt's result is not effective. There are some effective methods for solving Diophantine equations. One of the effective methods is based on estimates for linear forms in logarithms of algebraic numbers. Gelfand treated the case of linear forms in two logarithms following his work on Hilbert's seventh problem. In 1934, Gelfand – and independently Schneider – proved if α and β are algebraic then $\alpha \neq 0, 1$ and β is irrational then α^β is transcendental. In 1966, Baker extended this work to the case of linear forms in n logarithms of algebraic numbers with $n > 2$. As a consequence, he gave an effective procedure for solving Thue equations.

With Baker we gave a streamlined version of the argument to treat

$$x^3 - ay^3 = n. \tag{42}$$

Theorem 34 (Baker and Stewart)**.** *Let a and n be positive integers with a not a perfect cube. Then all solutions of in integers x and y satisfy*

$$\max(|x|, |y|) < (c_1 n)^{c_2}$$

where

$$c_1 = \varepsilon^{(50 \log\log \varepsilon)^2}$$
$$c_2 = 10^{12} \log \varepsilon,$$

and ε is the fundamental unit in the ring of algebraic integers of $\mathbb{Q}(\sqrt[3]{2})$.

Department of Pure Mathematics, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1

E-mail address: hsyang@uwaterloo.ca