Math 249A Fall 2010: Transcendental Number Theory

A course by Kannan Soundararajan
LaTeXed by Ian Petrow

September 19, 2011

Contents

1 Introduction; Transcendence of e and π

α is algebraic if there exists $p \in \mathbb{Z}[x]$, $p \neq 0$ with $p(\alpha) = 0$, otherwise α is called *transcendental* .

Cantor: Algebraic numbers are countable, so transcendental numbers exist, and are a measure 1 set in $[0, 1]$, but it is hard to prove transcendence for any particular number.

Examples of (proported) transcendental numbers: e, π, γ, e^π, $\sqrt{2}^{\sqrt{2}}$, $\zeta(3)$, $\zeta(5)\ldots$

Know: e, π, e^π, $\sqrt{2}^{\sqrt{2}}$ are transcendental. We don't even know if γ and $\zeta(5)$, $\zeta(7), \ldots$ are irrational or rational, and we know that $\zeta(3)$ is irrational, but not whether or not it is transcendental! Lioville showed that the number

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

is transcendental, and this was one of the first numbers proven to be transcendental.

Theorem 1 (Lioville)**.** *If $0 \neq p \in \mathbb{Z}[x]$ is of degree n, and α is a root of p, $\alpha \notin \mathbb{Q}$, then*

$$\left|\alpha - \frac{a}{q}\right| \geq \frac{C(\alpha)}{q^n}.$$

Proof. Assume without loss of generality that $\alpha < a/q$, and that p is irreducible. Then by the mean value theorem,

$$p(\alpha) - p(a/q) = (\alpha - a/q)p'(\xi)$$

for some point $\xi \in (\alpha, a/q)$. But $p(\alpha) = 0$ of course, and $p(a/q)$ is a rational number with denominator q^n. Thus

$$1/q^n \leq |\alpha - a/q| \sup_{x \in (\alpha-1, \alpha+1)} |p'(x)|.$$

□

This simple theorem immediately shows that Lioville's number is transcendental because it is approximated by a rational number far too well to be algebraic. But Lioville's theorem is pretty weak, and has been improved several times:

Theorem 2 (Thue)**.** *If $0 \neq p \in \mathbb{Z}[x]$ is of degree n, and α is a root of p, $\alpha \notin \mathbb{Q}$, then*

$$\left|\alpha - \frac{a}{q}\right| \geq \frac{C(\alpha, \varepsilon)}{q^{n/2+1+\varepsilon}},$$

where the constant involved is ineffective.

Theorem 3 (Roth)**.** *If α is algebraic, then*

$$\left|\alpha - \frac{a}{q}\right| \geq \frac{C(\alpha, \varepsilon)}{q^{2+\varepsilon}},$$

where the constant involved is ineffective.

Roth's theorem is the best possible result, because we have

Theorem 4 (Dirichlet's theorem on Diophantine Approximation)**.** *If $\alpha \notin \mathbb{Q}$, then $\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{q^2}$ for infinitely many q.*

Hermite: e is transcendental.
Lindemann: π is transcendental ($\therefore$ squaring the circle is impossible).
Weierstauß: Extended their results.

Theorem 5 (Lindemann)**.** *If $\alpha_1, \ldots, \alpha_n$ are distinct algebraic numbers, then $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are linearly independent over $\overline{\mathbb{Q}}$.*

Examples:

- Let $\alpha_1 = 0, \alpha_2 = 1$. This shows that e is transcendental.

- Let $\alpha_1 = 0, \alpha_2 = \pi i$. This shows that πi is transcendental.

Corollary 1. *If $\alpha_1, \ldots, \alpha_n$ are algebraic and linearly independent over $\mathbb{Q}$, then $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are linearly independent.*

Conjecture 1 (Schanuel's Conjecture)**.** *If $\alpha_1, \ldots, \alpha_n$ are any complex numbers linearly independent over $\mathbb{Q}$ then the transcendence degree of $\mathbb{Q}(\alpha_1, \ldots, \alpha_n, e^{\alpha_1}, \ldots, e^{\alpha_n})$ is at least n.*

If this conjecture is true, we can take $\alpha_1 = 1, \alpha_2 = \pi i$ to find that $\mathbb{Q}(\pi, e)$ has transcendence degree 2. This is an open problem!

Theorem 6 (Baker's Theorem)**.** *Let $\alpha_1, \ldots, \alpha_n$ be nonzero algebraic numbers. Then if $\log \alpha_1, \ldots, \log \alpha_n$ are linearly independent over $\mathbb{Q}$, then they're also linearly independent over $\overline{\mathbb{Q}}$.*

Exercise 1. *Show that e is irrational directly and quickly by considering the series for $n!e$.*

Claim 1. *e^n is irrational for every $n \in \mathbb{N}$.*

Proof. Let $f \in \mathbb{Z}[x]$. Define

$$I(u; f) = \int_0^u e^{u-t} f(t)\, dt = \int_0^u f(t)\, d(-e^{u-t}) = -f(u) + e^u f(0) + \int_0^u e^{u-t} f'(t)\, dt.$$

Iterating this computation gives

$$I(u; f) = e^u \sum_{j \geq 0} f^{(j)}(0) - \sum_{j \geq 0} f^{(j)}(u).$$

Note: f is a polynomial, so this is a finite sum. Now, assume e^n is rational. We derive a contradiction by finding conflicting upper and lower bounds for $I(n; f)$. The upper bound is easy:

$$|I(n; f)| \leq e^n \max_{x \in [0,n]} |f(x)|,$$

which grows like $C^{\deg f}$ in the f aspect. Now our aim is to try to find an f with $I(n; f) \geq (\deg f)!$ to contradict this upper bound. Pick $f(x) = x^{p-1}(x-n)^p$, where p is a large prime number. A short explicit computation shows

$$f^{(j)}(0) = \begin{cases} 0 & j \leq p-2 \\ (p-1)!(-n)^p & j = p-1 \\ \equiv 0 \pmod{p!} & j \geq p, \end{cases}$$

and

$$f^{(j)}(n) = \begin{cases} 0 & j \leq p-1 \\ \equiv 0 \pmod{p!} & j \geq p. \end{cases}$$

Assume p is large compared to n and the denominator of e^n. $I(n; f)$ is a rational integer divisible by $(p-1)!$ but not $p!$. So $|I(n; f)| \geq (p-1)!$. Contradiction. So e^n is irrational. □

Claim 2. *e is transcendental.*

Proof. Suppose not. Then $a_0 + a_1 e + \ldots + a_n e^n = 0$, $a_i \in \mathbb{Z}$. Define $I(u; f)$ as before. Then

$$I(u; f) = e^u \sum_{j \geq 0} f^{(j)}(0) - \sum_{j \geq 0} f^{(j)}(u)$$

and so

$$\sum_{k=0}^{n} a_k I(k, f) = -\sum_{k=0}^{n} a_k \sum_{j \geq 0} f^{(j)}(k).$$

Now choose $f(x) = x^{p-1}(x-1)^p \cdots (x-n)^p$. Similar to the above,

$$f^{(j)}(0) = \begin{cases} 0 & j \leq p-2 \\ (p-1)!(-1)^p \cdots (-n)^p & j = p-1 \\ \equiv 0 \pmod{p!} & j \geq p, \end{cases}$$

and for $1 \leq k \leq n$

$$f^{(j)}(k) = \begin{cases} 0 & j \leq p-1 \\ \equiv 0 \pmod{p!} & j \geq p. \end{cases}$$

Let p large compared to n and the coefficients $a_0, \ldots, a_n$. Then $I(n; f)$ is an integer divisible by $(p-1)!$ but not by $p!$, so $|I(n; f)| \geq (p-1)!$, but also $|I(n, f)| \leq C^p$ as before. Contradiction. □

Claim 3. *π is transcendental.*

Proof. Suppose not. Then πi is algebraic. Take $\alpha_1 = \pi i$, and let $\alpha_2, \ldots \alpha_n$ be all of the other Galois conjugates of πi. Then

$$(1 + e^{\alpha_1})(1 + e^{\alpha_2}) \cdots (1 + e^{\alpha_n}) = 0$$

because the first factor is 0. Expanding this, we get a sum of all possible terms of the form

$$\exp\left(\sum_{j=1}^{n} \epsilon_j \alpha_j\right)$$

where $\epsilon_j = 0, 1$. Some of these exponents are zero and some are not. Call the nonzero ones $\theta_1, \ldots, \theta_d$. Then we have

$$(2^n - d) + e^{\theta_1} + \cdots + e^{\theta_d} = 0.$$

As before, take our favorite auxiliary function,

$$I(u; f) = e^u \sum_{j \geq 0} f^{(j)}(0) - \sum_{j \geq 0} f^{(j)}(u).$$

Then we have

$$(2^n - d)I(0; f) + I(\theta_1; f) + \cdots + I(\theta_d; f) = -(2^n - d) \sum_{j \geq 0} f^{(j)}(0) - \sum_{k=1}^{d} \sum_{j \geq 0} f^{(j)}(\theta_k).$$

The right hand side of this expression is $\in \mathbb{Q}$ by Galois theory. Now take $A \in \mathbb{N}$ to clear the denominators of the α_j, i.e. so that $A\alpha_1, \ldots, A\alpha_n$ are all algebraic integers. Then let $f(x) = A^{dp}x^{p-1}(x-\theta_1)^p \cdots (x-\theta_d)^p$. As before,

$$f^{(j)}(0) \equiv \begin{cases} 0 \pmod{(p-1)!} & j = p-1 \\ 0 \pmod{p!} & \text{else,} \end{cases}$$

and $f^{(j)}(\theta_k)$ is always divisible by $p!$. So again, we have a that the right hand side of the above expression is an integer divisible by $(p-1)!$ but not by $p!$, so it is $\geq (p-1)!$ but also $\leq C^p$ by the same arguments as above. Contradiction. $\square$

Note the similarity of the last three proofs. We can generalize these, and will do so in the next lecture.

2 Lindemann-Weierstrauss theorem

Now we generalize the proofs of the transcendence of e and π from last time.

Theorem 7 (Lindemann-Weierstrauß)**.** *Let $\alpha_1, \ldots, \alpha_n$ be distinct algebraic numbers. Then*

$$\beta_1 e^{\alpha_1} + \cdots + \beta_n e^{\alpha_n} = 0$$

for algebraic $\beta_1, \cdots, \beta_n$ only if all $\beta_j = 0$. i.e. $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are linearly independent over $\overline{\mathbb{Q}}$.

This automatically gives us that e and π are transcendental, and proves a special case of Schanuel's conjecture.

Proof. Recall from the previous lecture that we defined

$$I(u,f) = \int_0^u e^{u-t} f(t)\, dt = e^u \sum_{j\geq 0} f^{(j)}(0) - \sum_{j \geq 0} f^{(j)}(u).$$

We will choose f to have a lot of zeros at integers or at algebraic numbers. We proceed as before, but things get a little more complicated. First, we make some simplifications.

First Simplification: All the β_j can be chosen to be rational integers. Why? Given a relation as in the theorem, we can produce another one with $\mathbb{Z}$ coefficients. We can consider

$$\prod_{\sigma \in Gal(\beta_1, \ldots, \beta_n)} (\sigma(\beta_1)e^{\alpha_1} + \cdots + \sigma(\beta_n)e^{\alpha_n})$$

instead. This expression is still 0 (one of its factors is zero), and upon expanding, it has rational coefficients. The expression for the each coefficient is a symmetric expression in the various σ, therefore fixed by $Gal(\beta_1, \ldots, \beta_n)$ and hence rational. We can then multiply through to clear denominators. If we show

that all the coefficients of this new expression are zero, it can only be because the original β_i were all zero (look at diagonal terms).

Second Simplification: We can take $\alpha_1, \ldots, \alpha_n$ to be a complete set of Galois conjugates. More specifically, we can assume that our expression is of the form

$$\beta_1 e^{\alpha_{n_0+1}} + \beta_1 e^{\alpha_{n_0+2}} + \cdots + \beta_1 e^{\alpha_{n_1}} + \beta_2 e^{\alpha_{n_1+1}} + \cdots + \beta_n e^{\alpha_{n_{t-1}+1}} + \cdots + \beta_n e^{\alpha_{n_t}},$$

where e.g. $\alpha_{n_0+1}, \ldots, \alpha_{n_1}$ is a complete set of Galois conjugates. Why can we do this? Take the original $\alpha_1, \ldots, \alpha_n$ to be roots of some big polynomial. Let $\alpha_{n+1}, \ldots, \alpha_N$ be the other roots of this polynomial. Take the product

$$\prod (\beta_1 e^{\alpha_{k_1}} + \cdots + \beta_n e^{\alpha_{k_n}})$$

where $\alpha_{k_1}, \ldots, \alpha_{k_n}$ are some choice of n of the $\{\alpha_i\}_{i=1}^N$, and the product is over all possible such choices. The original linear form is one of these, so this product equals 0. Expanding the product, we get a sum of terms of the form $e^{h_1\alpha_1 + \cdots + h_n\alpha_n}$, and if we do not simplify the coefficients $\beta_{n_1} \cdots \beta_{n_m}$, then the $h_1\alpha_{k_1} + \cdots + h_n\alpha_{k_n}$ which correspond to a given string of βs form a complete set of conjugates. Again, the only way that all of the coefficients of this expanded product are identically zero is if the original coefficients were all identically zero. This can be seen easily by estimating the size of the largest coefficient involved in this operation. So we are free to make these two simplifications.

We want to work with algebraic integers, but the $\alpha_1, \ldots, \alpha_n$ are a priori any algebraic numbers, so choose some large integer A which will clear all the denominators of the α_i. Then for every $1 \leq j \leq n$ we define

$$f_j(x) = \frac{A^{np}(x-\alpha_1)^p \cdots (x-\alpha_n)^p}{(x-\alpha_j)}.$$

This polynomial does not have $\mathbb{Z}$ coefficients but does have algebraic integer coefficients. We define

$$J_j = \sum_{k=1}^{n} \beta_k I(\alpha_k, f_j),$$

where p is a rational prime large compared to every other constant in the proof.

The plan is to show that $J_1 \cdots J_n \in \mathbb{Z}$ and that $J_1 \cdots J_n$ is divisible by $((p-1)!)^n$ but not by $p!$. Which implies $|J_1 \cdots J_n| \geq ((p-1)!)^n$, but then we also have $|J_1 \cdots J_n| \leq C^p$ by trivially estimating the integral defining $I(u, f)$, causing a contradiction and proving the theorem.

Folding the definition of $I(u, f)$ into that of J_j, we find

$$J_j = \sum_{k=1}^{n} \beta_k \left(e^{\alpha_k} \sum_{l \geq 0} f_j^{(l)}(0) - \sum_{l \geq 0} f_j^{(l)}(\alpha_k) \right) = -\sum_{k=1}^{n} \beta_k \sum_{l \geq 0} f_j^{(l)}(\alpha_k).$$

The second equality follows from the assumption $\sum \beta_k e^{\alpha_k} = 0$. Now we compute the derivatives of f_j. If $j \neq k$

$$f_j^{(l)}(\alpha_k) = \begin{cases} 0 & l \leq p-1 \\ \equiv 0 \pmod{p!} & l \geq p, \end{cases}$$

and if $j = k$

$$f_j^{(l)}(\alpha_k) = \begin{cases} 0 & l \le p-2 \\ A^{np}(p-1)! \prod_i (\alpha_k - \alpha_i)^p & l = p-1 \\ \equiv 0 \pmod{p!} & l \ge p. \end{cases}$$

Thus we see that each J_j is an algebraic integer divisible by $(p-1)!$ but not by $p!$ (using the first simplification).

Now we want to show $J_1 \cdots J_n \in \mathbb{Z}$. Using the second simplification above,

$$J_j = - \sum_{0 \le r \le t-1} \beta_{n_r+1} \sum_{k=n_r+1}^{n_{r+1}} \sum_{l \ge 0} f_j^{(l)}(\alpha_k).$$

The interior two sums is over a complete set of Galois conjugates α_k, so is Galois-invariant, hence in the ground field $\mathbb{Q}(\alpha_j)$. After taking the product $J_1 \cdots J_n$, we have an expression which is again Galois-invariant, hence $J_1 \cdots J_n \in \mathbb{Q}$. We assumed that A was large enough to cancel all denominators, and the first simplification says that the β are integers, hence $J_1 \cdots J_n$ is a rational integer. In fact, it is a rational integer divisible by $((p-1)!)^n$ but not by $p!$. We know that $I(u, f) \le C^{\deg f}$ in the f aspect, hence $|J_1 \cdots J_n| \le C'^p$ for some other constant C'. But this contradicts our lower bound! $\square$

But this proof seems totally unmotivated. How might one think of it? Well, if you want to prove that e is irrational, you can use the rapidly converging power series and truncate to obtain a simple proof (see exercise from previous lecture). Similarly, to show that e^z is irrational for algebraic z, one can use the power series

$$e^z = \sum_{j=0}^{N} \frac{z^j}{j!} + (\text{very small}).$$

This leads us to the idea of Padé approximations:

The idea is to find $B(z)$ and $A(z)$ so that $B(z)e^z - A(z)$ has many vanishing terms. We'll choose $B(z)$ to be a polynomial of degree L, say, and $A(z)$ to be degree M. We can choose the A and B so that the first $L + M$ terms vanish: write out the coefficients for A and B as $L+M$ unknowns. Then we get a system of $L + M$ equations in $L + M$ unknowns, so there is a solution. Therefore, it's possible to pick coefficients so that the first $L + M$ terms of $B(z)e^z - A(z)$ vanish.

Does this set-up seem familiar? It's exactly the same thing as our favorite interpolating function:

$$I(z, f) = \int_0^z e^{z-t} f(t)\, dt = e^z \sum_{j \ge 0} f^{(j)}(0) - \sum_{j \ge 0} f^{(j)}(z)$$

but now we think of f as depending on z. We might let $f(t)$ be something like $f(t) = t^M(z-t)^L$. Then $\sum f^{(j)}(0)$ and $\sum f^{(j)}(z)$ play the part of $B(z)$ and $A(z)$.

Now we move on to Baker's theorem. It is of fundamental importance in transcendence theory. For example, we have as a consequence the following result in diophantine approximation: If $0 \neq p \in \mathbb{Z}[x]$ is a polynomial of degree n, and $h = \text{height}(p) = \max|\text{coeffs}|$, then

$$|p(e)| \geq c(n, \epsilon)h^{-n-\epsilon}.$$

Theorem 8 (Baker's theorem on Linear forms in Logarithms)**.** *Let $\alpha_1, \ldots, \alpha_n$ be nonzero algebraic numbers. Assume that $\log \alpha_1, \ldots, \log \alpha_n$ are linearly independent over $\mathbb{Q}$. Then $1, \log \alpha_1, \ldots, \log \alpha_n$ are linearly independent over $\overline{\mathbb{Q}}$.*

Also, there is a quantitative version of this theorem, which we'll do later. Note also that the homogeneous version of this theorem, i.e. that $\log \alpha_1, \ldots, \log \alpha_n$ are linearly independent over $\overline{\mathbb{Q}}$, is slightly easier to prove. Baker's theorem generalizes the work of Gelfond and Schneider, who independently proved Hilbert's 7th problem in 1934: If α is algebraic, and β is an algebraic irrational, then α^β is transcendental. i.e. this is the case $n = 2$ of Baker's theorem: if $\log \alpha_1$ and $\log \alpha_2$ are linearly independent over $\mathbb{Q}$, then $\beta_1 \log \alpha_1 + \beta_2 \log \alpha_2 \neq 0$, $\beta_1, \beta_2 \in \overline{\mathbb{Q}}$, not both zero. Note also that in these problems you're allowed to pick any branch of the logarithm you like, so long as you (of course) stick to that one branch you've picked throughout the problem.

Baker's theorem has many beautiful Corollaries. For example, $e^\pi = (-1)^{-i} \notin \overline{\mathbb{Q}}$, and $\sqrt{2}^{\sqrt{2}} \notin \overline{\mathbb{Q}}$. More impressively,

Corollary 2. *Any $\overline{\mathbb{Q}}$ linear combination of logarithms of algebraic numbers is zero or transcendental.*

Proof. Suppose we have some $\alpha_1, \ldots, \alpha_n$ for which

$$\beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n = -\beta_0 \in \overline{\mathbb{Q}}.$$

Then if $\log \alpha_1, \ldots, \alpha_n$ are linearly independent over $\mathbb{Q}$ then we're done. Otherwise, $\log \alpha_n \in \text{span}(\log \alpha_1, \ldots, \log \alpha_{n-1})$. The corollary then follows by an induction argument on the dimension. □

Exercise: finish the details of this proof.

Corollary 3. *If $\beta_0, \beta_1, \ldots, \beta_n, \alpha_1, \ldots, \alpha_n$ are all not zero, then $e^{\beta_0}\alpha_1^{\beta_1}, \ldots, \alpha_n^{\beta_n}$ is transcendental.*

Proof. Exercise. □

Corollary 4. *If $1, \beta_1, \ldots, \beta_n$ are linearly independent over $\mathbb{Q}$, and $\alpha_i \neq 1 or\ 0$, then $\alpha_1^{\beta_1} \cdots \alpha_n^{\beta_n} \notin \overline{\mathbb{Q}}$.*

Proof. Exercise. □

Corollary 5. *$e^{\pi\alpha+\beta}$ is transcendental for all $\alpha, \beta \in \overline{\mathbb{Q}}$, not both zero. $\pi + \log \alpha \notin \overline{\mathbb{Q}}$ for any $0 \neq \alpha \in \overline{\mathbb{Q}}$.*

Proof. Exercise. □

3 Baker's Theorem, Part I

We take the rest of the week to prove Baker's Theorem, one of the most important theorems in Transcendence theory.

Theorem 9 (Baker's Theorem)**.** *Let $\alpha_1, \ldots, \alpha_n$ be algebraic and $\log \alpha_1, \ldots \log \alpha_n$ be linearly independent over $\mathbb{Q}$. Then for any $\beta_0, \ldots, \beta_n \in \overline{\mathbb{Q}}$ not all zero $\beta_0 + \beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n \neq 0$*

The homogeneous form (i.e. that $\beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n \neq 0$) is slightly easier. Baker's theorem is a generalization of

Theorem 10 (Gelford-Schneider)**.** *$\beta_1 \log \alpha_1 + \beta_2 \log \alpha_2 \neq 0$, so that $\alpha_2 \neq \alpha_1^{\beta_1}$, $\beta_1 \in \overline{\mathbb{Q}} \backslash \mathbb{Q}$, α_1, α_2 algebraic.*

This was Hilbert's seventh problem, which Hilbert thought would be solved after Fermat's last theorem and the Riemann Hypothesis.

Here's the plan:

1. A proposition involving an auxiliary function with various magical properties (the construction of which is the hardest part).
2. Why this implies Baker's theorem in the Homogeneous case
3. The construction of the auxiliary function for the case of Gelfond-Schneider
4. Return to the general case

Proof. We can assume without loss of generality that $\beta_n = -1$ Why? All the β_i, $i \geq 1$ can be zero, because then β_0 is 0 also. So we can divide through and change β_n to -1. Thus, we can assume there is a number $\alpha_n = e^{\beta_0} \alpha_1^{\beta_1} \cdots \alpha_{n-1}^{\beta_{n-1}} \in \overline{\mathbb{Q}}$, and try to derive a contradiction.

Let $h \in \mathbb{N}$ be large. Let $L = h^{2 - \frac{1}{4n}}$

Proposition 1. *There exists a function*

$$\phi(z) = \sum_{k_0, \ldots, k_n = 0}^{L} p(k_0, \ldots, k_n) z^{k_0} \alpha_1^{k_1 z} \ldots \alpha_n^{k_n z}$$

where $p(k_0, \ldots, k_n)$ are integers not all zero with size not too big. i.e. $|p(k_0, \ldots, k_n)| \leq \exp(h^3)$ and such that $\phi_j = \frac{d^j}{dz^j} \phi(z)$ and $\phi_j(0) \leq \exp(-h^{8n})$ for all $0 \leq j \leq h^{8n}$.

How many values of $p()$ are there? $(L+1)^{n+1}$ so something like h^{2n}. So the properties we want in such a function aren't completely trivial.

Now we go to the homogeneous case. We construct a similar

$$\phi(z) = \sum_{k_1, \ldots, k_n = 0}^{L} p(k_1, \ldots, k_n) \alpha_1^{k_1 z} \ldots \alpha_n^{k_n z}$$

with $|p(k_1,\dots,k_n)| \le \exp(h^3)$ and $\phi_j(0) \ll \exp(-h^{8n})$ for $j \le h^{8n}$. So what is this derivative?

$$\phi_j(0) = \sum_{k_1,\dots,k_n}^{L} p(k_1,\dots k_n)(k_1 \log\alpha_1 + \cdots + k_n \log\alpha_n)^j.$$

So there are $(L+1)^n - 1$ possible distinct values for the $(k_1 \log\alpha_1 + \cdots + k_n \log\alpha_n)$, which we call $\psi_1,\dots\psi_{R:=(L+1)^n-1}$. The distinctness follows from the assumed linear independence over $\mathbb{Q}$. We are going to show that the $\phi_j(0)$ are very very small. We'll think of the $p(k_1,\dots,k_n)$ as variables with coefficients ψ_i. So if we are to have these values very close to zero, we must have some serious approximate over-determination in the linear system above described. i.e. very small determinant. Actually, this is the Vandermonde determinant:

$$|\det\begin{pmatrix} 1 & \psi_1 & \psi_1^2 & \dots & \psi_1^{R-1} \\ \vdots & & \ddots & & \vdots \\ 1 & \psi_R & \cdots & & \psi_R^{R-1} \end{pmatrix}| = \prod_{j<k} |\psi_j - \psi_k|.$$

Now we do row operations on this determinant to make the first row $\phi_j(0)$. To do this, multiply the i-th row by the $p(k_1,\dots,k_n)$ which corresponds to that ψ_i and add it to the first row. Call the $p(k_1,\dots,k_n)$ corresponding to ψ_1 $p(l_1,\dots,l_n)$, say. Then the above determinant is

$$= \frac{1}{|p(l_1,\dots,l_n)|}|\det\begin{pmatrix} \phi_0(0) & \phi_1(0) & \phi_2(0) & \dots & \phi_{R-1}(0) \\ & \ddots & \ddots & & \\ 1 & \psi_i & \dots & & \psi_i^{R-1} \\ & & \ddots & \ddots & \end{pmatrix}|.$$

Then we expand this determinant along the top row, which as we previously remarked, will be shown to be very small. We'll show that each of the terms in the top row is $\ll \exp(-h^{8n})$, and we already know that each of the ψ_i are of size CL for some constant C. Then the above determinant is

$$\ll \exp(-h^{8n})(L^n)!(CL)^{L^{2n}},$$

where the factors come from the size of the terms in the first row, the number of terms, and the size of each of the ψ_i terms, respectfully. Recall that we set $L = h^{2-1/4n}$, so actually we have the above determinant is

$$\ll \exp\left(-\frac{h^{8n}}{2}\right).$$

But then we know that the original product for the Vandermonde determinant must be extremely small. There are L^{2n} factors in the Vandermonde determinant, so take the L^{2n}-th root. Thus for some $j < k$,

$$|\psi_j - \psi_k| \leq \exp\left(\frac{-h^{8n}}{L^{2n}}\right).$$

Great.

Next we have the following Lemma, which we will use to drive this estimate on the determinant to a contradiction.

Lemma 1. *If $k_1, \ldots, k_l$ are not all zero, $|k_1|, \ldots, |k_l| \leq L$ and $\alpha_1, \ldots, \alpha_n$ are algebraic with $\log \alpha_1, \ldots, \log \alpha_n$ linearly independent over $\mathbb{Q}$, then $|k_1 \log \alpha_1 + \cdots + k_n \log \alpha_n| \geq C^{-L}$ for some constant C.*

Proof. If $|k_1 \log \alpha_1 + \cdots + k_n \log \alpha_n|$ is small, then we can also say $|\alpha_1^{k_1} \cdots \alpha_n^{k_n} - 1|$ is small. Choose $A \in \mathbb{N}$ such that $A\alpha_j, A\alpha_j^{-1}$ are all algebraic integers. Then $A^{nL}(\alpha_1^{k_1} \cdots \alpha_n^{k_n} - 1)$ is an algebraic integer. So it's at least norm 1 if not 0. But because we assumed linear independence over $\mathbb{Q}$, it can only be 0 if $\alpha_1 \cdots \alpha_n$ is a root of unity. In this case, however, the lemma is trivially verified. So $N(A^{nL}(\alpha_1^{k_1} \cdots \alpha_n^{k_n} - 1)) \geq 1$. But it's also $\leq B^L |\alpha_1^{k_1} \cdots \alpha_n^{k_n} - 1|$. □

i.e. algebraic integers are norm at least 1, so they can't get very close together without being the same. The lemma now finishes the proof because $|\psi_j - \psi_k| \geq C^{-L}$, contradicting the bound just established.

Next we construct the auxiliary function in the $n = 2$ (Gelfond-Schneider) case. We assume that $\alpha_2 = \alpha_1^{\beta_1}$. Let h be large and let $L = h^{2-1/8}$. We are looking for $p(k_1, k_2)$ to define

$$\phi(z) = \sum_{k_1, k_2 = 0}^{L} p(k_1, k_2) \alpha_1^{k_1 z} \alpha_2^{k_2 z}.$$

We will take the $p(k_1, k_2) \in \mathbb{Z}$, not all zero, and such that $|p(k_1, k_2)| \leq \exp(h^3)$, and $|\phi_j(0)| \ll \exp(-h^{16})$ for $j \leq h^{16}$. So in this setting we are essentially trying to get h^{16} equations out of $h^{4-1/4}$ variables. How this is accomplished is really the magic of this proof. We have $h^{4-1/4}$ variables. First we solve h^3 equations, then magically solve all the other equations approximately by "lifting".

The first step is the choose $|p(k_1, k_2)| \leq \exp(h^3)$ such that $\phi_j(l) = 0$ for all $0 \leq j \leq h^2$, and $1 \leq l \leq h$. So that's about h^3 equations. Now we use the arithmetic data of the α_i.

$$\begin{aligned} \phi_j(l) &= \sum_{k_1, k_2 = 0}^{L} p(k_1, k_2)(k_1 \log \alpha_1 + k_2 \log \alpha_2)^j \alpha_1^{k_1 l} \alpha_2^{k_2 l} \\ &= (\log \alpha_1)^j \sum_{k_1, k_2 = 0}^{L} p(k_1, k_2)(k_1 + \beta_1 k_2)^j \alpha_1^{k_1 l} \alpha_2^{k_2 l} \end{aligned}$$

Now this last bit $(k_1+\beta_1 k_2)^j \alpha_1^{k_1 l}\alpha_2^{k_2 l}$ can be further reduced. Because $\alpha_1, \alpha_2, \beta_1$ are algebraic numbers, they satisfy polynomial relations. Thus large powers of any of these algebraic numbers can be reduced to linear combinations of smaller powers of them. In fact a linear combination of powers of $\alpha_1, \alpha_2, \beta_1$ smaller than the degree of each. So $(k_1+\beta_1 k_2)^j \alpha_1^{k_1 l}\alpha_2^{k_2 l}$ can be expressed as a linear combination of $\beta_1^{b_1}, \alpha_1^{a_1}, \alpha_2^{a_2}$, where $0 < b_1, a_1, a_2 \leq d-1$, where d is the maximum degree of $\alpha_1, \alpha_2, \beta_1$. (N.B. this will be explained more clearly in the next lecture). We then get

$$(k_1+\beta_1 k_2)^j \alpha_1^{k_1 l}\alpha_2^{k_2 l} = \sum_{a_1,a_2,b_1} \alpha_1^{a_1}\alpha_2^{a_2}\beta_1^{b_1} u(j,k_1,k_2,a_1,a_2,b_1)$$

for some coefficients $u(j,k_1,k_2,a_1,a_2,b_1)$ we get by applying the above described process. We can then set $\phi_j(l) = 0$ by solving d^3 linear equations

$$\sum p(k_1,k_2) u(j,k_1,k_2,a_1,a_2,b_1) = 0.$$

This should make you happy because there are $h^{4-1/4}$ choices for variables and d^3 (a constant) equations. Recall h can be chosen arbitrarily large.

Next we need to make $\phi(z)$ vanish at even more points. To do so, we will use the the following lemma.

Lemma 2 (Thue-Siegel)**.** *Suppose u_{ij}, $1 \leq i \leq M$, $1 \leq j \leq N$ are integers with $|u_{ij}| \leq U$. Want to solve*

$$\sum_{j=1}^{N} u_{ij} x_j = 0,$$

$x_1, \ldots, x_N \in \mathbb{Z}$, $N > M$. Then there is a nontrivial solution with $|x_j| \leq (NU)^{\frac{M}{N-M}}$.

Proof. Essentially, the Thue-Siegel lemma is a glorified version of the pigeonhole principle. Say $0 \leq x_j \leq X$. Then we have $(X+1)^N$ possibilities for the x_j. Consider the M-tuple $\{\sum_{j=1}^{N} u_{ij}x_j\}_{i=1,\ldots,M}$. So there are $(NUX)^M$ possible choices for this M-tuple in $\mathbb{Z}^M$. Then if $(X+1)^N > (NUX)^M$ there exists by the PHP a nontrivial solution to the system of equations. So there exists a solution with $|x_j| \leq (NU)^{\frac{M}{N-M}}$. □

4 Baker's Theorem, Part II

Recall that our goal was to prove Baker's *inhomogenous* theorem. In that theorem we are given $\alpha_1, \ldots, \alpha_n$ algebraic numbers, and $\log \alpha_1, \ldots, \log \alpha_n$ linearly independent over $\mathbb{Q}$. And we want to show $\beta_0 + \beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n \neq 0$ unless all $\beta_0, \ldots, \beta_n = 0$.

We also have the homogeneous version: $\beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n \neq 0$, and even easier, the Gelfond-Schneider result: $\beta_1 \log \alpha_1 + \beta_2 \log \alpha_2 \neq 0$. All of these

results follow from building some auxilliary function: assume there's a relation $\alpha_n = e_0^{\beta}\alpha_1^{\beta_1}\cdots\alpha_{n-1}^{\beta_{n-1}}$. This assumption is in place throughout the rest of the proof. Let h be a large integer, and $L := h^{2-1/4n}$. Then there is a function

$$\phi(z) = \sum_{k_0,\ldots,k_n}^{L} p(k_0,\ldots,k_n) z_0^{k_0}\alpha_1^{k_1 z}\cdots\alpha_n^{k_n z}$$

with the following properties:

1. $p(k_0,\ldots,k_n) \in \mathbb{Z}$
2. $|p(k_0,\ldots,k_n)| \leq \exp(h^3)$
3. $\phi_j(0) \ll \exp(-h^{8n})$ for all $j \leq h^{8n}$

To sumarize what we did last time: we considered the homogeneous version of this funciton

$$\sum_{k_1,\ldots,k_n}^{L} p(k_1,\ldots,k_n)\alpha_1^{k_1 z}\cdots\alpha_n^{k_n z}$$

and saw that computing a Vandermonde determinant led us to a contradiction. The idea is that Vandermonde forces two of the $k_1\log\alpha_1 + \cdots + k_n\log\alpha_n$ to be close together. But, as they're algebraic numbers, they can't be too close together, so in fact, they must actually be the same, and hence we get extra relations for free. Now focus on the Gelfond-Schneider case for simplicity. Suppose $\alpha_2 = \alpha_1^{\beta_1}$. The auxilliary funciton becomes

$$\phi(z) = \sum_{k_0,k_1}^{L} p(k_1,k_2)\alpha_1^{k_1 z}\alpha_2^{k_2 z}.$$

Aside: There's even a simpler version of this proof with $\alpha_1 = 2$, and $\alpha_2 = 3$ which will be presented in a subsequent lecture.

First step to find ϕ: Choose $p(k_1,k_2)$ such that $\phi_j(l) = 0$ for all $j \leq h^2$ and all $1 \leq l \leq h \in \mathbb{Z}$. For this, we'll need the

Lemma 3 (Thue-Siegel)**.** *Suppose we have M homogeneous linear equations in N variables, with $N > M$,*

$$\sum_{j=1}^{N} x_j u_{ij};$$

$j = 1,\ldots,M$, $|u_{ij}| \leq U$. Then there exists a nontrivial integer solution with $|x_j| \leq (2NU)^{\frac{M}{N-M}}$.

Here ends the summary of Monday's lecture.

We compute

$$\begin{aligned}\phi_j(l) &= \sum_{k_1,k_2=0}^{L} p(k_1,k_2)(k_1\log\alpha_1 + k_2\log\alpha_2)^j \alpha_1^{k_1 l}\alpha_2^{k_2 l} \\ &= (\log\alpha_1)^j \sum_{k_1,k_2=0}^{L} p(k_1,k_2)(k_1+k_2\beta_1)^j \alpha_1^{k_1 l}\alpha_2^{k_2 l}\end{aligned}$$

where $\alpha_1, \alpha_2, \beta_1$ are algebraic of degree at most d. We can express $\beta^j \alpha_1^{k_1 l}\alpha_2^{k_2 l}$ as a linear combination of terms of the form $\beta_1^{b_1}\alpha_1^{a_1}\alpha_2^{a_2}$. We might as well clear demoninators to make $\phi(z)$ a combination of algebraic integers. So assume without loss of generality that the $\alpha_i$are algebraic integers. So suppose $\alpha_1^d = A_0 + A_1\alpha_1 + \cdots + A_{d-1}\alpha_1^{d-1}$ be the defining relation for α_1. Let $\mathrm{ht}(\alpha_1) = \max_{1\le i\le d-1}|A_i|$. Also, if we multiply again by α_1, we get further relations like $\alpha_1^{d+1} = A_0\alpha_1 + A_1 + \cdots + A_{d-1}\alpha_1^d$, to which we can apply the relation for α_1^d again. Thus we find that the coefficients of α_1^{d+1} are $\le 2(\mathrm{ht}(\alpha_1))^2$. In this way I can control the size of the coefficients of any α_1^j.

Now for $j \le h^2$ and $l \le h$, we want

$$B^{h^2+2LH}\sum_{k_1,k_2=0}^{L} p(k_1,k_2)(k_1+k_2\beta_1)^j \alpha_1^{k_1 l}\alpha_2^{k_2 l} = 0$$

where we have chosen B so that all denominators are cleared. We expand

$$B^{h^2+2LH}(k_1+k_2\beta_1)^j\alpha_1^{k_1 l}\alpha_2^{k_2 l} \ll (100L)^{h^2}CLh \ll \exp h^3,$$

where the first term comes from the binomial coefficients, and the second from the heights. The B factor can easily be absorbed into the other factors. Now we want to apply the Thue-Siegel lemma. In terms of the statement of that lemma we take $N = (L+1)^2$, and the u_{ij} to be $(k_1+k_2\beta_1)^j\alpha_1^{k_1 l}\alpha_2^{k_2 l}$. The x_j will be the $p(k_1,k_2)$. So $U = \exp(h^3)$, and there are $M = d^3h^3$ equations. Recall $L = h^{2-1/8}$, so that $N = h^{4-1/4}$. So by Thue-Siegel, we get a nontrivial solution for the $p(k_1,k_2)$ which is $\le (2h^{4-1/4}\exp(h^3))^{\frac{d^3h^3}{h^{4-1/4}-d^3h^3}} \le \exp(h^3)$.

So, we've constructed a

$$\phi(z) = \sum p(k_1,k_2)\alpha_1^{k_1 z}\alpha_2^{k_2 z}$$

with $\phi_j(l) = 0$ for $j \le h^2$, $|p(k_1,k_2)| \ll \exp(h^3)$ and $l \le h$. This is still pretty far off from the auxiliary function we promised at the beginning of the proof. Baker's idea is that we can get even more vanishing out of this function. We can push the function to get $\phi_j(l) = 0$ for $j \le h^2/2$ and $l \le h^{1+1/8n}$. i.e. Reduce the order of vanishing but increase the number of distinct zeros. We have that the complex function

$$\frac{\phi_j(z)}{((z-1)(z-2)\cdots(z-h))^{h^2/2}}$$

is holomorphic in the entire complex plane. Let $R > 2h$ be some number to be chosen later, and let $R > r > h \in \mathbb{N}$. Then we apply the maximum modulus principle:

$$\frac{\phi_j(r)}{((r-1)(r-2)\cdots(r-h))^{h^2/2}} \leq \max_{|z|=R} \frac{|\phi_j(z)|}{|(z-1)\cdots(z-h)|^{h^2/2}}$$

So we get the bound

$$|\phi_j(r)| \leq r^{h^3/2} \left(\frac{R}{2}\right)^{-h^3/2} \exp(h^3 + Ch^2 \log L + cRL)$$

where the first factor comes from the denominator of the LHS, the second from the denominator of the RHS, and the third from three factors in each term of $\sum p(k_1,k_2)(k_1 \log \alpha_1 + k_2 \log \alpha_2)\alpha_1^{k_1 z}\alpha_2^{k_2 z}$. We now take R as to minimize this bound, and find that $R = \frac{h^3}{2CL} \approx h^{1+1/8}$. So taking $r \leq h^{1+1/16}$ is reasonable. Hence we obtain

$$|\phi_j(r)| \ll \exp(-ch^3 \log h + Ch^3)$$

with $|R/r| \geq h^{1/16}$. So we get that $\phi_j(r)$ is very small, but why is it actually zero? Consider now

$$B^{2Lr+j}\phi_j(r) = (\log \alpha_1)^j B^{2Lr+j} \sum_{k_1+k_2} p(k_1,k_2)(k_1 + \beta_1 k_2)^j \alpha_1^{k_1 r}\alpha_2^{k_2 r}.$$

Omitting the first factor on the RHS, we have an algebraic integer (choosing B large enough to cancel the denominators of $\alpha_1, \alpha_2, \beta_1$). We can also say that this algebraic integer lies in a field of degree at most d^3. Furthermore, all of its conjugates are $\leq \exp(h^3 + cRL + Ch^2 \log L) \leq \exp(2h^3)$, by the same calculation as above. An algebraic integer has norm ≥ 1 if it is not zero. So if it is not zero, $|\phi_j(r)| \geq \exp(-c'h^3)$, using the fact that $p(k_1,k_2) \in \mathbb{Z}$. But this contradicts the bound we got from the maximum modulus principle! Thus these additional $\phi_j(r)$ must actually be zero.

So what happens if we take $j \leq h^2/4$, $l \leq h^{1+2/16}$ and $\phi_j(l) = 0$ and try to do the same thing? We consider the function

$$\frac{\phi_j(z)}{((z-1)(z-2)\cdots(z-\lfloor h^{1+1/16} \rfloor))^{h^2/4}},$$

which is holomorphic in the entire complex plane with $h^{1+1/16} < r$ and $R > 2h^{1+1/8}$. So by the same arguments as above,

$$|\phi_j(r)| \leq r^{\frac{h^2}{4}h^{1+1/16}} \left(\frac{R}{2}\right)^{\frac{h^2}{4}h^{1+1/16}} \exp(h^3 + Ch^2 \log L + cRL)$$

as before. The choice for R that optimizes this is then $R = \frac{h^2 h^{1+1/16}}{CL} = h^{1+1/8+1/16}$, and we can take $r \leq h^{1+1/16+1/16}$.

So observe that we get a constant increase in the admissible range of r by $\frac{1}{16}$ every time we decrease the range of j by a factor of two. So, repeating this process, we can take, say $j \leq \frac{h^2}{2^{256}}$ and get $\phi_j(r) = 0$ for $r \leq h^{16}$. We now want to show that $\phi_j(0)$ is very small. So consider

$$\frac{\phi_j(z)}{((z-1)(z-2)\cdots(z-h^{16}))^{h^2/2^{256}}},$$

which is entire. We want an estimate for $|\phi(z)|$ on the circle $|z| = 1$. So consider a large circle of radius $R > 2h^{16}$, then by the same maximum modulus trick as above, we get that for z on $|z| = 1$

$$|\phi(z)| \leq (h^{16})^{\frac{h^2}{2^{256}}h^{16}} \left(\frac{R}{2}\right)^{-\frac{h^2}{2^{256}}h^{16}} \exp(Ch^3 + cRL).$$

To minimize this, we take $R = \frac{h^{18}}{2^{256}CL} \approx h^{16+1/8}$. So we get that $|\phi(z)| \ll \exp(-h^{18})$. Now, by the Cauchy integral formula,

$$\phi_j(0) = \frac{j!}{2\pi i}\int_{|z|=1} \frac{\phi(z)}{z^{j+1}}\, dz \ll j! \exp(-h^{18}).$$

If $j \leq h^{16}$ then $\phi_j(0) \ll \exp(-h^{16})$. This finishes the construction of our auxiliary function.

Thus, the Vandermonde calculation and Lemma 1 from Lecture 3 are valid, and produce a contradiction, which proves the theorem.

□

Next time we will generalize this to the inhomogeneous case, and the case of n variables instead of just $n = 2$. After that, we will show a simple proof for $\alpha_1 = 2$ and $\alpha_2 = 3$.

5 Baker's Theorem, Part III

Today we tackle the general case of Baker's theorem. Let $\alpha_1, \ldots, \alpha_n$ be algebraic numbers with $\log \alpha_1, \ldots, \log \alpha_n$ linearly independent over $\mathbb{Q}$. Assume $\beta_0, \ldots, \beta_n \in \overline{\mathbb{Q}}$ are not all zero, and by dividing through we can also assume without loss of generality that $\beta_n = -1$. So we have and will use the relation $\alpha_n = e^{\beta_0}\alpha_1^{\beta_1}\cdots\alpha_{n-1}^{\beta_{n-1}}$. The main additional difficulty is the construction of the auxiliary function. We are looking for a function

$$\phi(z) = \sum_{k_0,\ldots,k_n=0}^{L} p(k_0, \ldots, k_n) z^{k_0} \alpha_1^{k_1 z}\alpha_2^{k_2 z}\cdots\alpha_n^{k_n z}$$

with $p(k_0, \ldots, k_n) \in \mathbb{Z}$, $|p(k_0, \ldots, k_n)| \ll \exp(h^3)$, and $\phi_j(0) \ll exp(-h^{8n})$ for $j \leq h^{8n}$. We will deduce Baker's inhomogeneous theorem from the existence

of such a function, but first, let's concentrate on the construction of such a function.

We want to do the same trick from last time where we pull out a $(\log \alpha_1)^j$ and find an algebraic number. But we can't do that here because there are many different $\log \alpha_i$ factors. So we need to introduce a function of several variables, which will allow us to do essentially the same thing. Let $L = h^{2-1/4n}$. Here is how to define the function:

$$\Phi(z_0, \dots, z_{n-1}) = \sum_{k_0,\dots,k_n=0}^{L} p(k_0, \dots, k_n) z_0^{k_0} e^{k_n\beta_0 z_0} \alpha_1^{(k_1+k_n\beta_1)z_1} \alpha_2^{(k_2+k_n\beta_2)z_2} \cdots \alpha_{n-1}^{(k_{n-1}+k_n\beta_{n-1})z_{n-1}}.$$

Now, when we take partial derivatives in the various variables, we still get the algebraic property which we desire. Let $\phi(z) := \Phi(z, \dots, z)$. Let

$$\Phi_{m_0,\dots,m_{n-1}}(z_0, \dots, z_{n-1}) := \frac{\partial^{m_0}}{\partial z_0^{m_0}} \frac{\partial^{m_1}}{\partial z_1^{m_1}} \cdots \frac{\partial^{m_{n-1}}}{\partial z_{n-1}^{m_{n-1}}} \Phi(z_0, \dots, z_{n-1}).$$

Let

$$\phi_j(z) := \sum_{m_0+\cdots+m_{n-1}=j} \binom{j}{m_0, \dots, m_{n-1}} \Phi_{m_0,\dots,m_{n-1}}(z, \dots, z).$$

So we'll make this Φ by demanding that for all choices of $m_0 + \cdots + m_{n-1} \leq h^2$, and for all $l \leq h$ we have $\Phi_{m_0,\dots,m_{n-1}}(l, \dots, l) = 0$. On first inspection, this seems feasible, as we have $h^{2n+1}d^{2n}$ equations to satisfy and L^{n+1} total variables.

$$\Phi_{m_0,\dots,m_{n-1}}(l, \dots, l) = \sum_{k_0,\dots,k_n=0}^{L} p(k_0, \dots, k_n) \frac{\partial^{m_0}}{\partial z_0^{m_0}} (z_0^{k_0} e^{k_n\beta_0 z_0})|_{z_0=l} (k_1+k_n\beta_1)^{m_1}$$

$$\times\ \alpha_1^{(k_1+k_n\beta_1)l} \cdots (k_{n-1}+k_n\beta_{n-1})^{m_{n-1}} \alpha_{n-1}^{(k_{n-1}+k_n\beta_{n-1})l} (\log \alpha_1)^{m_1} (\log \alpha_2)^{m_2} \cdots (\log \alpha_{n-1})^{m_{n-1}}.$$

So each summand in the above is (up to the p's and the $\log \alpha$'s)

$$\alpha_1^{k_1 l} \alpha_2^{k_2 l} \cdots \alpha_n^{k_n l} q(k_0, \dots, k_n, \beta_0, \dots, \beta_{n-1}),$$

where we reduce using the relation $\alpha_n = e^{\beta_0} \alpha_1^{\beta_1} \cdots \alpha_{n-1}^{\beta_{n-1}}$, and the $q(\cdots)$ is some polynomial. Now we can reduce this even further using the polynomial relations which define each algebraic number involved. Thus we can express it in terms of a polynomial combination of terms of the form $\alpha_1^{a_1}, \alpha_2^{a_2}, \dots, \alpha_n^{a_n} \beta_0^{b_0} \cdots \beta_{n-1}^{b_{n-1}}$ with $0 \leq a_j, b_j \leq d-1$. Finally, we also want to cancel denominators. The $m_0, \dots, m_n \leq h^2$, so the coefficients required will be of size $B^{h^2+nLh}(CL)^h$ times an expression in terms of the heights of the α_i, β_i, which can be absorbed into the B term. This whole thing is $\ll \exp(h^3)$.

Now we are in a position to apply the Thue-Siegel lemma. We have $d^{2n}h^{2n+1}$ equations, L^{n+1} free variables and the size of the coefficients is $\ll \exp(h^3)$. Thus by the Thue-Siegel lemma, we have a nontrivial solution for the $p(k_0,\dots,k_n)$ with

$$\begin{aligned} |p(k_0,\dots,k_n)| &\leq (L^{n+1}\exp(h^3))^{\frac{d^{2n}h^{2n+1}}{L^{n+1}-d^{2n}h^{2n+1}}} \\ &\ll \exp(h^3) \end{aligned}$$

So we have quite a bit of vanishing, but we need even more!

Extrapolation Step: If $m_0+\cdots+m_{n-1} \leq \frac{h^2}{2}$ and $r \leq h^{1+1/8n}$ then $\Phi_{m_0,\dots,m_{n-1}}(r,\dots,r)=0$. Take

$$f(z) := \Phi_{m_0,\dots,m_{n-1}}(z,\dots,z),$$

with $j \leq h^2/2$. And also

$$f_j(z) = \sum_{j_0+\cdots+j_{n-1}=j} \binom{j}{j_0,\dots,j_{n-1}} \Phi_{m_0+j_0,\dots,m_{n-1}+j_{n-1}}(z,\dots,z)$$

so $f_j(z)=0$ for $j \leq h^2/2$ and $z=l \leq h$. Now consider

$$\frac{f(z)}{((z-1)\cdots(z-h))^{\frac{h^2}{2}}},$$

which is an entire function. Choose $R>2h$, and $r>h$. We apply the maximum modulus principle just like last week's lecture. So we get that $|f(r)| \leq r^{h^3/2}\left(\frac{R}{2}\right)^{-h^3/2}\max_{|z|=R}|f(z)| = r^{h^3/2}\left(\frac{R}{2}\right)^{-h^3/2}\exp(2h^3+CLR)$. If we optimize, we find that it's best to take $R \approx \frac{h^3}{CL} \approx h^{1+1/4n}$, so if $r \leq h^{1+1/8n}$, we conclude that

$$|f_j(r)| \leq \exp(-ch^3\log h).$$

(N.B. The use of the maximum modulus theorem above is not essential to the proof. It is but a crutch. We'll do this proof yet again in yet another way which avoids the use of the maximum modulus principle. Probably next lecture.)

As we already discussed, $f(r)$ suitably multiplied is an algebraic integer, all of whose conjugates are $\exp(h^3)$, and degree d^{2n}. But then, as in previous lectures, by computing norms we find that $f(r)=0$. Thus we've proven the extrapolation step.

By iterating the extrapolation step, we can, for any $s \in \mathbb{N}$, take $m_0+\cdots+m_{n-1} \leq \frac{h^2}{2^s}$, and $r \leq h^{1+s/8n}$. In particular, taking $s=(8n)^2$,we get a Φ such that for any $m_0+\cdots+m_{n-1} \leq \frac{h^2}{2^{(8n)^2}}$, and $r \leq h^{1+8n}$

$$\Phi_{m_0,\dots,m_{n-1}}(r,\dots,r)=0.$$

Putting $\phi(z)=\Phi(z,\dots,z)$, we get $\phi_j(r)=0$ for $j \leq \frac{h^2}{2^{(8n)^2}}$ and $r \leq h^{1+8n}$.

Now we finish the construction of the auxiliary function using the Cauchy integral formula in similar fashion to the previous lecture. Consider

$$\frac{\phi(z)}{((z-1)\cdots(z-h^{1+8n}))^{\frac{h^2}{2^{(8n)^2}}}}.$$

So if $|z| \leq 1$ and some $R > 2h^{1+8n}$ then

$$|\phi(z)| \leq ((h^{1+8n})!)^{h^2} \left(\frac{R}{2}\right)^{-\frac{h^{3+8n}}{(\cdots)}} \exp(Ch^3 + cLR).$$

Then optimizing, we set $R = \frac{h^{3+8n}}{CL}$. So then

$$\phi_j(0) = \frac{j!}{2\pi i} \int_{|z|=1} \frac{\phi(z)}{z^{j+1}}\, dz \ll j! \exp(-c' h^{3+8n} \log h),$$

for $j \leq h^{8n}$, $\phi_j(0) \ll \exp(-h^{3+8n})$.

Thus we've finished the construction of the auxiliary function. We still need to deduce Baker's theorem, but nothing much changes there from the previous instances of the proof, so we'll do it quickly next time.

So what are some of the key points of the proof?

1. The norm of an algebraic integer is 0 or ≥ 1. If $|\alpha|$ is tiny and its conjugates $|\sigma(\alpha)|$ are not huge, then the algebraic integer must be 0.
2. Solving linear equations (Thue-Siegel)
3. Extrapolation argument

Jeff Lagarias' comment: Where do we use the fact that Φ is a multivariable function? It is in the extrapolation step that we crucially use that the partial derivatives go in many directions instead of just along the diagonal. The multivariate nature of Φ seems to be essential, even though we don't use any complex analysis of several variables or anything like that.

6 Baker Concluded; Powers of 2 and 3

Last time we wrote down the auxiliary function for the inhomogeneous Baker's theorem:

$$\Phi(z_0, \ldots, z_{n-1}) = \sum_{k_0,\ldots,k_n=0}^{L} p(k_0, \ldots, k_n) z_0^{k_n} e^{k_n \beta_0 z_0} \alpha_1^{(k_1 + k_n\beta_1)z_1} \cdots \alpha_{n-1}^{(k_{n-1} + k_n \beta_{n-1})z_{n-1}}$$

and also that

$$\phi(z) = \Phi(z, z, \ldots, z) = \sum_{k_0,\ldots,k_n=0} p(k_0, \ldots, k_n) z^{k_0} \alpha_1^{k_1 z} \cdots \alpha_n^{k_n z}$$

with $\phi_j(0) \ll \exp(-h^{8n})$ for all $j \leq h^{8n}$. We now actually deduce Baker's theorem. We first go back to the homogeneous case.

$$\phi(z) = \sum_{k_0,\ldots,k_n=0} p(k_0,\ldots,k_n)\alpha_1^{k_1 z}\cdots\alpha_n^{k_n z};$$

$$\phi_j(0) = \sum p(k_0,\ldots,k_n)\left(\sum_{i=1}^{n} k_i \log\alpha_i\right)^j.$$

Let $\psi_1,\ldots,\psi_{(L+1)^n}$ be the values of $\sum k_i\log\alpha_i$. Then we have

$$\sum_{r=1}^{(L+1)^n} p(r)\psi_r^j \ll \exp(-h^{8n})$$

for $r \leq h^{8n}$. One of these coefficients is nonzero. Say $p(\tilde{r})$. Choose a polynomial W with $W(\psi_{\tilde{r}}) = 1$ and $W(\psi_r) = 0$ for $r \neq \tilde{r}$. Write the coefficients of W:

$$W(z) = \sum_{j=0}^{(L+1)^n-1} w_j z^j.$$

Then

$$\begin{aligned} p(\tilde{r}) &= \sum_{r=1}^{(L+1)^n} p(r)W(\psi_r) \\ &= \sum_{j=0}^{(L+1)^n-1} w_j \sum_{r=1}^{(L+1)^n} p(r)\psi_r^j \\ &= \sum_{j=0}^{(L+1)^n-1} w_j\phi_j(0) \end{aligned}$$

We know all the $\phi_j(0)$ are tiny. If the w_j are not too big, then we'll be done, because $p(\tilde{r}) \geq 1$. What would be a good definition for W? We can take

$$W(z) = \prod_{r\neq\tilde{r}} \frac{(z-\psi_r)}{(\psi_{\tilde{r}}-\psi_r)}.$$

Recall that $|\psi_r - \psi_s| \gg \exp(-CL)$ to estimate the coefficients w_j.

Now we go back to the Inhomogeneous case. Assume that $\psi_1,\ldots\psi_{(L+1)^n}$ are distinct values of $\sum k_i\log\alpha_i$. Then

$$\phi(z) = \sum_{k_0=0}^{L}\sum_{r=1}^{(L+1)^n} p(k_0,r)z^{k_0}e^{\psi_r z}.$$

We know that we can rig up $\phi_j(0)$ as an $(L+1)^{n+1} \times (L+1)^{n+1}$ determinant which is nonsingular and all entries small. Pick out a $\tilde{k}_0, \tilde{r}$ with $p(\tilde{k}_0, \tilde{r}) \neq 0$. Find W a polynomial such that $W_j(\psi_r) = 0$ if $r \neq \tilde{r}$ and $j \leq L$. And

$$W_j(\psi_{\tilde{r}}) = \begin{cases} 0 & \text{if } j \neq \tilde{k}_0 \\ 1 & \text{if } j = \tilde{k}_0 \end{cases}$$

The subscript j means derivative, as it is used with ϕ. Let $W(z) = \sum_{j=1}^{(L+1)^n} w_j z^j$. Computing the derivatives by hand we have $W_{k_0}(\psi_r) = \sum_j w_j j(j-1)\cdots(j-k_0+1)\psi_r^{j-k_0}$. As in the previous calculation, we have

$$\begin{aligned} p(\tilde{k}_0, \tilde{r}) &= \sum_{k,r} p(k,r) W_{k_0}(\psi_r) \\ &= \sum_{j=0}^{(L+1)^n - 1} w_j \sum_{k_0, r} p(k_0, r) j(j-1)\cdots(j-k_0+1)\psi_r^{j-k_0}. \end{aligned}$$

But

$$j(j-1)\cdots(j-k_0+1)\psi_r^{j-k_0} = \frac{d^j}{dz^j}(z^{k_0} e^{\psi_r z})|_{z=0}.$$

So we have

$$p(\tilde{k}_0, \tilde{r}) = \sum_j w_j \phi_j(0).$$

Now, we are done by the same principle as before. Let's write down exactly what W is.

$$W(z) = \prod_{r \neq \tilde{r}} \left(\frac{z - \psi_r}{\psi_{\tilde{r}} - \psi_r} \right)^{L+1} \frac{(z-\psi_{\tilde{r}})^{k_0}}{k_0!} (1 + a_1(z-\psi_{\tilde{r}}) + a_2(z-\psi_{\tilde{r}})^2 + \cdots + a_{L+1-k_0}(z-\psi_{\tilde{r}})^{L+1-k_0}),$$

then solve for the $a_1, a_2, \ldots$. That the ψ_r are well-spaced implies that the w_j are not too huge. In fact of size about

$$\left(\frac{c \max |\psi_r|}{\min_{r \neq s} |\psi_r - \psi_s|} \right)^{L^{n+1}}.$$

But this is like size $\exp(h^{2n})$, which loses to $\exp(-h^{8n})$. So we're done.

Plan: We'll go through this proof again and try to prove a quantitative lower bound. Many wonderful theorems follow from effective estimates which we get out of Baker's theorem, as we shall see. Most other theorems in transcendence theory are ineffective.

Here's the problem: Let $S = \{p_1, \ldots, p_s\}$. The the S-integers are $p_1^{\alpha_1}, \ldots, p_s^{\alpha_s}$. Then put these in order: $n_1 < n_2 < \ldots$. The question is, how close can n_i and n_{i+1} get? Up to X there are about $\sim C(\log X)^s$ numbers. To answer this question, there is the following

Theorem 11 (Tijdeman)**.** *There exists a constant $C = C(S)$ such that*

$$n_{i+1} - n_i \gg \frac{n_i}{(\log n_i)^{C(S)}},$$

where the suppressed constants are effective.

Corollary 6.

$$|2^m - 3^n| \gg \frac{2^m}{m^C}.$$

Effectively.

We'll prove this corollary, but not Tijdeman's theorem. In fact, we won't even prove the full strength of the corollary. We hope to show

$$|2^m - 3^n| \geq \frac{2^m}{\exp((\log m)^C)}$$

for some $0 < C < 1$. We do it by taking Gelfond-Schneider for $\alpha_1 = 2$, $\alpha_2 = 3$. We knew $\frac{\log 2}{\log 3}$ was irrational, now we know it's transcendental. Suppose that $\frac{\log 2}{\log 3} = \frac{U}{V} + \delta$. Goal: bound δ. This is the same as $2^V = 3^U + \delta V$; so $|2^V - 3^U| \asymp V\delta 3^U$. Assume δ is very small. (We'll eventually get better than $\delta \geq 3^{-U}/V$).

Plan: Examine proof of Gelfond-Schneider.

$$\phi(z) = \sum_{k_1,k_2=0}^{L} p(k_1,k_2) 2^{k_1 z} 3^{k_2 z}.$$

We want to construct something of this type with various properties. (Eventually, we'll take L to be a tiny power of $\log U$ or $\log V$, so L is small compared to the rational approximation).

$$\begin{aligned}
\phi_j(z) &= \sum_{k_1,k_2=0}^{L} p(k_1,k_2)(k_1 \log 2 + k_2 \log 3)^j 2^{k_1 z} 3^{k_2 z} \\
&= (\log 3)^j \sum_{k_1,k_2=0}^{L} p(k_1,k_2) \left(k_1 \left(\frac{U}{V} + \delta \right) + k_2 \right) 2^{k_1 z} 3^{k_2 z}
\end{aligned}$$

Let

$$\tilde{\phi}(z,j) = (\log 3)^j \sum_{k_1,k_2=0}^{L} p(k_1,k_2) \left(\frac{k_1 U}{V} + k_2 \right)^j 2^{k_1 z} 3^{k_2 z}.$$

$\tilde{\phi}(z,j)$ is $\frac{(\log 3)^j}{V^j}$ times an integer for $z \in \mathbb{N}$. Suppose $|p(k_1,k_2)| \leq P$. Then

$$\phi_j(z) = \tilde{\phi}(z,j) + O(\delta (2L)^{j+3} 6^{L|z|} P),$$

where in the error term, the 2^j comes from $(\log 3)^j$, the L^{j+1} comes from the binomial coefficients, and the extra L^2 comes from the L^2 terms. We want to

solve $\tilde{\phi}(z,j)=0$ for $1\le z=r\le R_0$, and $j\le J_0$. We'll take R_0 the be a bit less than h^2, say $L^{1-2\alpha}$, and J_0 to be about $L^{1+\alpha}$. The number of equations is then R_0J_0 and the number of free variables is L^2. The coefficients of the R_0J_0 equations are $\le (L(U+V))^j 6^{LR_0}$. Thus we can use the Thue-Siegel lemma. We get a nontrivial solution with

$$P\le \left(3L^2(L(U+V))^J 6^{LR_0}\right)^{\frac{J_0R_0}{L^2-J_0R_0}}.$$

7 Effective Baker for $\log 2$ and $\log 3$

We continue the proof from last time about powers of 2 and 3. Let's review briefly what happened last time. We let

$$\frac{\log 2}{\log 3}=\frac{U}{V}+\delta,$$

with δ very small. This is equivalent to

$$2^V=3^{U+\delta V}=3^U+O(\delta V 3^U).$$

Tijdeman's theorem (which is a consequence of effective estimates in Baker's theorem, and which we won't prove) would give $|\delta|\ge V^{-c}$ for some constant c. We will prove

Theorem 12.

$$\delta\gg \exp(-c_\kappa(\log V)^\kappa)$$

for any $\kappa>2$.

Tijdeman's theorem would give the above as a corollary with $\kappa=1$. So then we got started:

$$\phi(z)=\sum_{k_1,k_2=0}^{L} p(k_1,k_2)2^{k_1z}3^{k_2z}$$

We'll let L be a power of $\log V$, and that power will be >1. Next compute the derivatives

$$\phi_j(z)=\sum_{k_1,k_2}^{L} p(k_1,k_2)(k_1\log 2+k_2\log 3)^j 2^{k_1z}3^{k_2z}.$$

Assume that $|p(k_1,k_2)|\le P$. We'll establish a bound for P later. We have $k_1\log 2+k_2\log 3=(\log 3)(k_1(\frac{U}{V}+\delta)+k_2)$. Let

$$\phi(z,j)=(\log 3)^j\sum_{k_1,k_2} p(k_1,k_2)(\frac{k_1U}{V}+k_2)^j 2^{k_1z}3^{k_2z}$$

so that

$$\phi_j(z)=\phi(z,j)+O(\delta P 6^{L|z|}(2L)^{j+2}).$$

Here ends the summary of the previous lecture.

The algebraic input to this demonstration is via this $\phi(z,j)$. Indeed, if $z \in \mathbb{N}$, then $\phi(z,j) = (\text{integer}) \left(\frac{\log 3}{V}\right)^j$. Now, the construction of such a function is in terms of integers, so we start as usual by Thue-Siegel. To make $\phi(z,j) = 0$ for all $j \leq J_0$, $z = r \leq R_0$, we must have $J_0 R_0 \leq L^2$. Later, we'll take $J_0 = L^{1+\alpha}$, and $R_0 = L^{1-2\alpha}$. We'll take α very small, eventually like $\kappa - 2$.

Now we figure out the quantities we'll need for Thue-Siegel. We have $\frac{k_1 U}{V} + k_2 \approx 2LV$, so that the size of the coefficients is

$$\left(\frac{k_1 U}{V} + k_2\right)^{J_0} 6^{LR_0}.$$

The number of variables involved is L^2, so Thue-Siegel implies that there is a nontrivial solution for the $p(k_1, k_2)$ of size

$$\leq \left((2LV)^{J_0} 6^{LR_0}\right)^{\frac{(1.001)J_0 R_0}{L^2}} = \exp(3\frac{J_0^2 R_0}{L^2} \log V + 2\frac{J_0 R_0^2}{L})$$

so that we take $P := \exp(3L \log V + 2\frac{J_0 R_0^2}{L}) = \exp(3L \log V + 2L^{2-3\alpha})$.

Now we move on to the extrapolation step. $j \leq J_0$, $r \leq R_0$, so that we have (forgetting the +2's etc...)

$$\phi_j(z) = O(\delta P 6^{LR_0} (2L)^{J_0}).$$

Now we extrapolate to bound $\phi_j(r)$ with $j \leq J_1$, $r \leq R_1 = R_0 L^{\alpha/2} = L^{1-2\alpha+\alpha/2}$. Recall that the extrapolation step was the key point in the proof of Baker's theorem! Before, we used the maximum modulus principle, but we can't use that here because we don't have that something actually equals zero, but is only very small. But we can get around this. Define as before

$$\frac{\phi_j(z)}{((z-1)\cdots(z-R_0))^{J_0/2}},$$

with $r > R$, $R > 2r > 2R_0$. We'll also perform the same integration

$$\frac{1}{2\pi i} \int_{|z|=R} \frac{\phi_j(z)}{((z-1)\cdots(z-R_0))^{J_0/2}} \frac{dz}{(z-r)}$$

but we no longer know that the integrand is entire. If it were so, this would have been the Cauchy integral formula. We estimate this integral in two different ways, and compare the results.

First: Bound the integral trivially along the circle. Each factor in the denominator is at least $(R/2)^{-R_0 J_0/2}$, and the numerator will be $\max_{|z|=R} |\phi_j(z)|$, so that the whole integral is

$$\ll \left(\frac{R}{2}\right)^{-\frac{R_0 J_0}{2}} P(2L)^{\frac{J_0}{2}} 6^{LR}.$$

Choose $R \asymp R_0 J_0/L = L^{\alpha/2} R_1$.

Second:

$$\frac{\phi_j(r)}{((r-1)\cdots(r-R_0))^{J_0/2}} + \sum_{1 \le k \le R_0} Res_{k=z} \left(\frac{\phi_j(z)}{((z-1)\cdots(z-R_0))^{J_0/2}(z-r)} \right)$$

This residue calculation might look horrible to you, but it's really not as bad as it seems, and we have to do it. The worst case will be when $z = R_0/2$, and using the estimate $R_0! \approx 10^{J_0/2}(R_0/2)!$, I claim that what you get is at most

$$\ll \frac{(10J_0)^{\frac{J_0}{2}}}{(R_0!)^{\frac{J_0}{2}}} \delta P \exp(2LR_0 + J_0 \log L)$$

$$\ll \delta P \exp(2LR_0 + 2J_0 \log J_0 - \frac{R_0 J_0}{2} \log \frac{R_0}{2}).$$

So we have that

$$|\phi_j(r)| \ll (R_1)^{\frac{R_0 J_0}{2}} ((R/2)^{-\frac{R_0 J_0}{2}} P(2L)^{\frac{J_0}{2}} 6^{LR} + \delta P)$$

At this point, we don't know anything about this δP term. Further simplifying:

$$\ll P(2L)^{\frac{J_0}{2}} 6^{LR} \exp(-\frac{R_0 J_0}{3} \log^{\alpha/2} L) + \delta P \exp(\frac{R_0 J_0}{2} \log R_1).$$

Now, at this point, if the second term is large, then δ must be large, but then we'd be done. But on the other hand, the first term is smaller than anything.

What is the analogue of the norm step? $\phi(z, j)$ is small, but by integrality properties, it will turn out to actually be zero. Then we will need to extrapolate. If $z \in \mathbb{N}$, then $\phi(z, j) = (\text{integer}) \left(\frac{\log 3}{V}\right)^j$. So if our bound for this is $\le V^{-j}$, then in fact $\phi(z, j) = 0$. It suffices to show the following

1. $\delta \ll \exp(-J_0 \log V - R_0 J_0 \log L)$. If this is false, then we'd be done, because we already have a lower bound for δ.
2. From the "smaller than anything" term above, it suffices to show that $\alpha R_0 \log L \gg \log V$.

So if these are satisfied, we use that we know

$$\phi_j(r) = O(\delta P 6^{Lr} (2L)^j)$$

for $r \le R_1$, $j \le J_1$ to go to $j \le J_1/2 = J_2$ and $r \le R_2 = R_1 L^{\alpha/2}$. So we can run through the same argument and get the same thing at the last step. This is possible if $\delta \le \exp(-J_0 \log V - R_1 J_1 \log R_2)$ and $\alpha R_0 \log L > \log V$. Do this k times, assuming it's possible to do so. If it's not possible, it's for a good reason, because δ is too big, and we stop. So at the end we have

$$\phi_j(r) = O(\delta P 6^{LR_k} (2L)^{J_k})$$

for $j \leq J_k = \frac{J_0}{2^k}$ and $r \leq R_k = R_0 L^{k\alpha/2}$. Deduce $\phi_j(0)$ is small for many values of j.

The next step is to estimate $\phi(w)$ for $|w| = 1/2$. Then

$$\phi_j(0) = \frac{j!}{2\pi i} \int_{|w|=1/2} \frac{\phi(w)}{w^{j+1}}\, dw.$$

Consider now

$$\frac{1}{2\pi i} \int_{|z|=R} \frac{\phi(z)}{((z-1)\cdots(z-R_k))^{J_k}} \frac{dw}{(z-w)}$$

then bound it in two different ways, as before.

What you get if you do this:

$$\phi(w) \ll P 6^{LR} \left(\frac{2R_k}{R}\right)^{R_k J_k} + \delta \exp(2L \log V + R_k J_k \log R).$$

Now we choose R. A non-optimal but sufficient choice is $R = R_k L^{\alpha/2}$. Thus we get

$$\ll P \exp(-\frac{\alpha R_k J_k}{5} \log L) + \delta \exp(2L \log V + R_k J_k \log R_{k+1})).$$

And as a consequence

$$\phi_j(0) \ll j^j \left(P \exp(-\frac{R_k J_k}{5} \log L) + \delta \exp(2L \log V + R_k J_k \log R_{k+1}) \right).$$

Great. Now, what was it that we wanted? Recall

$$\phi_j(0) = \sum_{k_1,k_2=0}^{L} p(k_1,k_2)(k_1 \log 2 + k_2 \log 3)^j.$$

Let $\psi_1, \ldots, \psi_{(L+1)^2}$ be the distinct values of $(k_1 \log 2 + k_2 \log 3)$.

$$\phi_j(0) = \sum_{r=1}^{(L+1)^2} p(r)\psi_r^j$$

so if this is small, there is some $\psi_r - \psi_s$ is very small. So we want some estimate for all $j \leq (L+1)^2 + 1$. We write down the Lagrange interpolation formula in the homogenous case. First of all there is some $\tilde{r}$ so that $p(\tilde{r}) \neq 0$, so that we take

$$W(z) := \prod_{r \neq \tilde{r}} \frac{(z-\psi_r)}{(\psi_{\tilde{r}} - \psi_r)} = \sum_{j=0}^{(L+1)^2-1} w_j z^j.$$

Then

$$p(\tilde{r}) = \sum p(r) W(\psi_r) = \sum_{j=0}^{(L+1)^2-1} w_j \phi_j(0),$$

so we want to show that the w_j are not too big to finish. We want to show the denominators are well-spaced to bound the coefficients. We use the stupid but sufficient bound $|a \log 2 + b \log 3| \geq 1/2^a$ to get the denominator in

$$w_j \ll \frac{(2L)^{L^2}}{2^{L^2}} \ll \exp(3L^2 \log L).$$

We want that for $j \leq (L+1)^2$, $\phi_j(0) \ll \exp(-4L^2 \log L)$. So if

$$\phi_j(0) \ll P \exp(-\frac{\alpha R_k J_k}{5} \log L) + \delta \exp(2L \log V + R_k J_k \log R_{k+1})) \ll \exp(-10L^2 \log L),$$

say, then we get a contradiction. Assume that $\delta \ll \exp(-10^6 L^{2+\alpha} \log L)$ to get this contradiction. The we can take $R_k J_k = L^{2+\alpha}$, and $L^{1-2\alpha} > \log V$ or equivalently $L > (\log V)^{1+3\alpha}$ to get the bound and the contradiction. The false assumption was that $\delta \ll \exp(cL^{2+\alpha} \log L)$, so then $\delta \gg \exp(-c(\log V)^{2+\alpha})$, and we're done.

So, we've now done about the first 20 pages in Baker's book. It's very dense. Next time we'll do some applications of this effective result. e.g. the class number one problem.

8 Applications: Class Number One

So proving the theorem about separation of powers of 2 and 3 wasn't that easy after all, but we did it. We've proved

$$|2^V - 3^U| \gg_\kappa \frac{2^V}{\exp((\log V)^\kappa)},$$

for any $\kappa > 2$. Now, you can believe that we can do this in general for many bases. The corresponding bound for linear forms in logarithms is as follows: Let $\beta_i \in \overline{\mathbb{Q}}$, and $\log \alpha_1, \ldots, \log \alpha_n$ be linearly independent over $\mathbb{Q}$. Let $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ all have degree $\leq d$. Let $\mathrm{ht}(\beta_0), \ldots, \mathrm{ht}(\beta_n) \leq B$, and $\mathrm{ht}(\alpha_0), \ldots, \mathrm{ht}(\alpha_n) \leq A$. Then

$$|\beta_0 + \beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n| \gg \exp(-C(\log B)^\kappa),$$

where $\kappa > n+1$, and $C = C(\kappa, n, d, A)$. In the homogeneous case, we get better, $\kappa > n$. This is the main result from Baker's landmark 1968 paper in Mathematika.

But this isn't the best possible result. A better result was proven by Feldman, in which we are allowed to take $\kappa = 1$, obtaining a bound like $\gg B^{-C}$, with $C = C_1(\log A)^{k_1}$. We can take, $k_1 = n$, say. Even from here, various refinements and improvements are possible. The state of the art is contained in a paper by Baker and Wristholz which appears in Crelle's Journal sometime in

the 90s. Morally, Feldman's result is some sort of effective power savings over Lioville's theorem. From the same viewpoint, Baker's theorem would have given

$$\frac{1}{q^d} \exp((\log q)^{1/k}).$$

This is not as good as Roth's theorem, but goes beyond Lioville, and unlike most results in this field, is effective, which can be used to great consequence.

So there are lot of applications of these results. The first one we'll do is the class number one problem. Class number one is a very old problem, going all the way back to Gauss. It asks one to compute all immaginary quadratic fields with class number one: $-1, -3, \ldots, -163$. There are 9 of them in all. The first result in the direction of this problem was due to Heilbronn.

Theorem 13 (Heilbronn)**.** *$h(-d) \to \infty$ as $d \to \infty$*

The proof is famous for splitting into two cases, first assuming that GRH is true, which is the easier case, and secondly, assuming GRH is false. It uses a purported exceptional zero $L(\rho, \chi) = 0$ and controls everything else in terms of this zero. But of course Heilbronn's result is completely ineffective because we can't find such a zero.

Next we have

Theorem 14 (Landau-Siegel)**.** *$h(-d) \geq C(\epsilon) d^{1/2-\epsilon}$*

But $C(\epsilon)$ above is completely ineffective. None of these results rule out the 10th possible imaginary quadratic field with class number one.

Heegner in 1952 solved class number one, but his work was ignored for some time. But then Stark solved class number one in 1968 using techniques from diophantine equations. Later Stark filled in the unproven statements in Heegner's proof and showed that it, in fact, was correct. At the same time, Baker also proved class number one, using a method of Gelfond and Linnik from 1949. Gelfond and Linnik actually got extremely close to to proving class number one but got unlucky. With one additional trick, their proof would work to prove class number one using Gelfond and Schneider's 1934 transcendence result. They thought they needed linear independence of three logarithms, but actually only two would have been sufficient.

Jeff Lagarias' comment: Actually, to prove class number one, you need the work of both Baker and Stark. Stark proved that there was no 10th field between -163 and a very very large but finite number, and Baker proved that there was no imaginary quadratic field with class number one and discriminant very very large.

In 1971 Baker and Stark proved that there were no class number two fields with discriminant larger than 10^{1000}, or some number like that. We won't prove this. Baker's work does not seem to prove the class number problem effectively in general, but this is known due to work of Goldfeld, and Gross-Zagier.

We now begin the proof of the class number one problem, i.e. that there is an effective upper bound to the discriminant of a field with class number one.

Claim 4. *Suppose $\mathbb{Q}(\sqrt{-d}) = 1$. Then if $p \leq \frac{d+1}{4}$, and $\chi = \left(\frac{-d}{\cdot}\right)$ is primitive, then $\chi(p) = -1$, which is the same as saying that $x^2 + x + \frac{d+1}{4}$ takes only prime values in $0 \leq x \leq \frac{d+1}{4}$.*

Proof. Suppose not Then $(p) = \mathfrak{p}\overline{\mathfrak{p}}$. Then we have

$$N(\mathfrak{p}) = p = N\left(\frac{x + y\sqrt{-d}}{2}\right) = \frac{x^2 + dy^2}{4},$$

so if not zero, $p \geq \frac{d+1}{4}$. But we assumed otherwise. Contradiction. □

Proof (Class Number One). Let $K = \mathbb{Q}(\sqrt{-d})$.

$$\zeta_K(s) = \zeta(s)L(s, \chi_{-d}) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} \left(1 - \frac{\chi_{-d}(p)}{p^s}\right)^{-1} = \zeta(2s) \prod_{p > \frac{d+1}{4}} \frac{1 + \frac{1}{p^s}}{\left(1 - \frac{\chi_{-d}(p)}{p^s}\right)}.$$

But we really don't want to work with a pole, so instead, let's take $\psi \mod q$ to be a real quadratic character, e.g. the one associated to the real quadratic field $\mathbb{Q}(\sqrt{21})$. Then we have

$$\begin{aligned} & L(s, \psi)L(s, \psi\chi_{-d}) \\ = & \prod_p \left(1 - \frac{\psi(p)}{p^s}\right)^{-1} \left(1 - \frac{\psi(p)\chi_{-d}(p)}{p^s}\right)^{-1} \\ = & \zeta(2s) \prod_{p|q} \left(1 - \frac{1}{p^{2s}}\right) + \sum_{n > \frac{d+1}{4}} \frac{a(n)}{n^s} \end{aligned}$$

for some choice of $a(n)$ obeying a bound like $a(n) \leq d(n)$. Now, compute using the left hand side. Take the completed L-function:

$$\Lambda(s) = \left(\frac{q}{\pi}\right)^{s/2} \Gamma(s/2)L(s, \psi) \left(\frac{dq}{\pi}\right)^{s/2} \Gamma\left(\frac{s+1}{2}\right) L(s, \psi\chi_{-d}).$$

The nice feature of having one odd and one even character is the form of the gamma factors which appear here. We can use the duplication formula! So this

$$= (2\sqrt{\pi}) \left(\frac{q^2 d}{4\pi^2}\right)^{s/2} \Gamma(s)L(s, \psi)L(s, \psi\chi_{-d}) = \Lambda(1 - s).$$

The idea is to get a nice formula for $\Lambda(1)$ with extreme precision. Let $c > 1$.

Then

$$\begin{aligned}
&\frac{1}{2\pi i}\int_{(c)}\Lambda(s)\left(\frac{1}{s-1}+\frac{1}{s}\right)ds\\
=\;&\Lambda(1)+\Lambda(0)+\frac{1}{2\pi i}\int_{(1-c)}\Lambda(s)\left(\frac{1}{s-1}+\frac{1}{s}\right)ds\\
=\;&2\Lambda(1)+\frac{1}{2\pi i}\int_{(1-c)}\Lambda(1-s)\left(\frac{1}{s-1}+\frac{1}{s}\right)ds\\
=\;&2\Lambda(1)-\frac{1}{2\pi i}\int_{(c)}\Lambda(w)\left(\frac{1}{1-w}+\frac{1}{-w}\right)(-dw)
\end{aligned}$$

Which gives

$$\begin{aligned}
\Lambda(1) &= \frac{1}{2\pi i}\int_{(c)}\Lambda(s)\frac{2s-1}{s(s-1)}\,ds\\
&= \frac{2\sqrt{\pi}}{2\pi i}\int_{(c)}\left(\frac{q\sqrt{d}}{2\pi}\right)^s\Gamma(s)\left(\zeta(2s)\prod_{p|q}\left(1-\frac{1}{p^{2s}}\right)+\sum_{n>\frac{d+1}{4}}\frac{a(n)}{n^s}\right)\frac{2s-1}{s(s-1)}\,ds.
\end{aligned}$$

Now, I claim that the term in the above coming from the dirichlet series makes an extremely small contribution. Pulling out this term

$$\sum_{n>\frac{d+1}{4}}a(n)\frac{1}{2\pi i}\int_{(c)}\Gamma(s)\frac{2s-1}{s(s-1)}\left(\frac{q\sqrt{d}}{2\pi n}\right)^s ds,$$

we shift the contour to get an estimate on it's size. One of the factors looks like $\sqrt{d}^c$ and $\Gamma(s)$ looks like $\left(\frac{c}{e}\right)^c$ by Stirling's formula, so we choose c to balance these. The minimum of $\Gamma(c)\xi^c$ is at about $\xi = c$, so that the above is $\ll \exp(-C\frac{2\pi n}{q\sqrt{d}})$, where the C only comes from adjusting for the $2s - 1/s(s-1)$ factor. Because our range of summation is over $n > (d+1)/4$, this whole term is actually $\exp(-a\sqrt{d}/q)$ for some $a > 0$. So it's not even a good estimate in the d aspect, it's an extremely good estimate!

Now, the main term is

$$\frac{2\sqrt{\pi}}{2\pi i}\int_{(c)}\left(\frac{q\sqrt{d}}{2\pi}\right)^s\Gamma(s)\zeta(2s)\prod_{p|q}\left(1-\frac{1}{p^{2s}}\right)\left(\frac{2s-1}{s(s-1)}\right)ds.$$

Move the contour to the left, to $-c$. How far? Enough to balance the size of the two terms $\left(\frac{q\sqrt{d}}{2\pi}\right)^s$ and $\Gamma(s)\zeta(2s)$. Again, we pick up an extremely small error term, the main term coming from the residues. So, the main term above is

$$\text{Residues} + O(\exp\left(\frac{-a\sqrt{d}}{q}\right)).$$

So what are these residues? There is no pole at $2s = 1$ because it is canceled by a zero. There is a pole at $s = 1$, whose residue we must account for. There is also a possible double pole at $s = 0$ coming from $\Gamma(s)/s$. But there are zeros at $s = 0$ coming from

$$\prod_{p|q} \left(1 - \frac{1}{p^{2s}}\right),$$

and if q has at least two distinct prime factors, then we have a double zero, and hence no pole at zero!

So taking the residue at $s = 1$, we have found that

$$\begin{aligned} \Lambda(1) &= 2\sqrt{\pi}\frac{q\sqrt{d}}{2\pi}\zeta(2)\prod_{p|q}\left(1 - \frac{1}{p^2}\right) + O(\exp\left(\frac{-a\sqrt{d}}{q}\right)) \\ &= 2\sqrt{\pi}\frac{q\sqrt{d}}{2\pi}L(1,\psi)L(1,\psi\chi_{-d}), \end{aligned}$$

by which we obtain

$$L(1,\psi)L(1,\psi\chi_{-d}) = \zeta(2)\prod_{p|q}\left(1 - \frac{1}{p^2}\right) + O(\exp\left(\frac{-a\sqrt{d}}{q}\right)).$$

Now we're really close to finished.

$$L(1,\psi) = \frac{\log \epsilon_q h(q)}{\sqrt{q}}; \qquad L(1,\psi\chi_{-d}) = \frac{h(-dq)\pi}{\sqrt{qd}}$$

so that

$$\frac{h(q)h(-qd)\log\epsilon_q}{q\sqrt{d}} = \frac{\pi}{6}\prod_{p|q}\left(1 - \frac{1}{p^2}\right) + O(\exp\left(\frac{-a\sqrt{d}}{q}\right))$$

and

$$6h(q)h(-qd)q\log\epsilon_q = \pi\sqrt{d}\prod_{p|q}(p^2-1) + O(\exp\left(\frac{-a\sqrt{d}}{q}\right)).$$

$$\pi = -i\log(-1),$$

so this is a contradiction to Baker's theorem, which gives a bound of $\exp(-(\log d)^3)$. So d is in fact bounded. □

This proof is very special to the case of class number one. Let's look at what happens for class number two. $h(-d) = 2$. Genus theory says that the number of times which 2 divides $h(-d)$ depends only on the number of prime factors of d. Let $d = p_1q_1$, with $p_1 \equiv 1 \pmod 4$, and $q_1 \equiv 3 \pmod 4$. Let χ be

a character mod d given by the Legendre symbol. We still have that all small primes $p < \frac{d+1}{4}$ are inert, except for p_1 and q_1 which are ramified. Then

$$L(s,\psi)L(s,\psi\chi) = \zeta(2s)\prod_{p|q}\left(1-\frac{1}{p^{2s}}\right)\left(1-\frac{\psi(p_1)\left(\frac{p_1}{q_1}\right)}{p_1^s}\right)^{-1}\left(1-\frac{\psi(q_1)\left(\frac{q_1}{p_1}\right)}{q_1^s}\right)^{-1}.$$

Now, before, we had $\left(\frac{q\sqrt{d}}{2\pi}\right)^{-s}$, but now we have both p_1 and q_1, with $p_1q_1 \approx d$, with one of these small and one large with respect to $\sqrt{d}$. If p_1 and q_1 are far apart, we can make the same argument. The hard case is $p_1 \approx q_1 \approx \sqrt{d}$. Then we have error of $\exp(-\sqrt{d}/p)$, and this destroys our whole argument. So instead we now take

$$L(s,\psi\chi_{p_1})L(s,\psi\chi_{p_2}) + L(s,\psi)L(s,\psi\chi)$$

and things will cancel out in a nice way to make things work. But we'll talk about this next time.

9 Applications: Class Number Two, a Putnam Problem, the Unit Equation

Another way to think about the proof of class number one which generalizes more easily to higher class number problems is via Eisenstein series. Consider the series

$$E^*(z,s) = \zeta(2s)E(z,s) = \sum_{(c,d)\neq(0,0)}\frac{y^s}{|cz+d|^{2s}},$$

where

$$E(z,s) = \sum_{\gamma\in\Gamma_\infty\backslash SL_2(\mathbb{Z})}(\mathrm{Im}\gamma z)^s.$$

Let $\mathcal{Q}_{-D}$ denote the set of positive definite integral binary quadratic forms of discriminant $-D$. Let $\mathcal{Q}_{-D}/\Gamma$ denote the set of equivalence classes of such quadratic forms under the usual change of basis action. We define $h(-D) = |\mathcal{Q}_{-D}/\Gamma|$. To each such equivalence class there is associated a unique point $z = \frac{-b+\sqrt{-D}}{2a} \in \Gamma\backslash\mathcal{H}$ called the Heegner point of discriminant $-D$. We denote the set of Heegner points of discriminant $-D$ as Λ_{-D}. Then if $[a,b,c]$ is a quadratic form corresponding to the Heegner point z_0, then

$$E^*(z_0,s) = \left(\frac{\sqrt{D}}{2}\right)^s\sum_{(x,y)\neq(0,0)}\frac{1}{(ax^2+bxy+cy^2)^s} = \left(\frac{\sqrt{D}}{2}\right)^s\sum_{n=1}^{\infty}\frac{r_{[a,b,c]}(n)}{n^s},$$

where $r_Q(n)$ is the representation number of n by the quadratic form $Q \in \mathcal{Q}_{-D}$. Observe that

$$\sum_{Q \in \mathcal{Q}_{-D}} \frac{r_Q(n)}{|Aut(Q)|} = \sum_{k|n} \left(\frac{-D}{k}\right),$$

whence

$$\sum_{z \in \Lambda_{-D}} \frac{E^*(z,s)}{|Aut(Q_z)|} = \left(\frac{\sqrt{D}}{2}\right)^s \sum_{n=1}^{\infty} \sum_{k|n} \left(\frac{-D}{k}\right) n^{-s} = \left(\frac{\sqrt{D}}{2}\right)^s \zeta(s) L(s, \chi_{-D}) = \left(\frac{\sqrt{D}}{2}\right)^s \zeta_{\mathbb{Q}(-D)}(s).$$

Again, we don't want to work with the pole here, but that's okay, just as in last week's lecture, we can twist by a character to remove it. This costs us a twist in the Eisenstein series by the same character, which will raise the level. But it does so by a bounded and controllable amount, so such a twist is pretty benign. We will again like to get an exponentially good approximation to $L(1,\psi)L(1,\psi\chi_{-D})$, and derive a contradiction with Baker's theorem. To get the approximation, we look at the Fourier expansion of the Eisenstein series. $E^*(z,s)$ has a Fourier expansion of the form

$$E^*(z,s) = (\text{const}) + \sum_{n \neq 0} (\cdots) e(nx) W_s(yn)$$

Where $W_s(\cdot)$ is some sort of exponentially decreasing Bessel function which depends on s. If y is sufficiently large, the exponential decay of W_s kicks in and the series drops off exponentially fast in n. If we assume that $h(-D) = 1$, then we only have one Eisenstein series to deal with, and we know that the one Heegner point comes from the quadratic form $x^2 + x + \frac{d+1}{4}$, hence $a = 1$ so we know that the imaginary part of the Heegner point is $\sqrt{D}/2$, which is sufficiently large to take advantage of the exponential decay of W_s. So truncating the Fourier expansion, we get an exponentially good approximation to the central value, which leads to a contradiction with Baker's theorem, as last time, which implies that D must be bounded.

Now, what happens to this method when we try to do class number two? We instead get two Heegner points. If both of these points are high in the cusp, the above method works. However, this need not be so. The best general bound on a we have is $a \leq \frac{\sqrt{d}}{\sqrt{3}}$, which just says that the Heegner point lies in the fundamental domain. So that's not really that helpful. However, to solve class number two, we can use genus characters. Most hopeful in this direction is a problem of Euler, "Numeri Idonei". i.e. Find all d for which there is only one class per genus. To approach this problem one is led to consider

$$\sum_{d_1 d_2 = d} L(s, \psi\chi_{d_1}) L(s, \psi\chi_{d_2}),$$

which actually puts another kink in our method because we get many logarithms and their heights start to grow very fast. The situation is manageable for class number 2, but not for class number 4. There's no hope at all for class number 3 because we don't even have any genus theory at our disposal.

It's also important to mention the results of Goldfeld, Gross-Zagier, who prove an effective class number bound like $h(-d) \gg \log |d|$, which is the best effective bound know, but still very very far from the truth.

But a few final remarks on genus theory: The number of genera is something like $2^{\omega(d)}$, where $\omega(d)$ is the number of prime factors of d. It would also be possible to attack one class per genus with Goldfeld/ Gross-Zagier. A lower bound of any power of d, e.g. $h(-d) \gg d^{.0000000001}$ would be sufficient to solve one class per genus, but we don't even have this!

Okay. So we're now done with the class number problem. Let's move on to other applications of effective Baker.

Application 2: Here's a very Putnam-esqe application: $1, 3, 8, 120$ have the property that if you multiply any two and add 1, then you get a square.

Question 1. *Are there any $n > 120$ for which $1, 3, 8, n$ have the same property?*

These sorts of sets are called "Diophantine tuples". The answer is "NO", as solved by Baker and Davenport in a 1968 paper. We'll prove this result, and see that it's not actually as isolated a result as it at first seems. We want to solve the system of equations

$$\begin{aligned} n+1 &= \square \\ 3n+1 &= \square \\ 8n+1 &= \square \end{aligned}$$

Which is just the same as solving

$$\begin{aligned} n &= x^2 - 1 \\ 3x^2 - 2 &= y^2 \\ 8x^2 - 7 &= z^2 \end{aligned}$$

So we are really trying to solve effectively the hyperelliptic equation

$$t^2 = (3x^2 - 2)(8x^2 - 7),$$

and by solve effectively I claim that there are only finitely many solutions to such an equation, and we can write down an explicit bound for how large they may be. Siegel showed that *ineffectively* that there are only finitely many solutions in integers, but Baker's theorem will solve the problem effectively. We can re-write our equations as

$$y^2 - 3x^2 = -2; \qquad z^2 - 8x^2 = -7,$$

and these are simultaneous pell-type equations. $(2 + \sqrt{3})$ is a unit in $\mathbb{Q}(\sqrt{3})$, so for each $n \in \mathbb{Z}$ there exists y_n and x_n for which

$$y + \sqrt{3}x = (y_n + \sqrt{3}x_n)(2 + \sqrt{3})^n.$$

So, we get a lot of solutions

$$y_n^2 - 3x_n^2 = -2.$$

We can assume also that

$$1 < y_n + \sqrt{3}x_n < 2 + \sqrt{3}$$

i.e.

$$\frac{2}{2+\sqrt{3}} < \sqrt{3}x_n - y_n < 2$$

so we want to solve

$$1 + \frac{2}{2+\sqrt{3}} \leq 2\sqrt{3}x_n \leq 4 + \sqrt{3},$$

which implies $x_n = 1$, forcing $y_n = 1$. So we're found there's a finite list (i.e. 1) of solutions. now we do the same thing to the next equation. For each $m \in \mathbb{Z}$ we have a solution to

$$(z + \sqrt{8}x) = (z_m + \sqrt{8}x_m)(3 + \sqrt{8})^m.$$

The same process gives two choices:

$$(\pm 1 + \sqrt{8})(3 + \sqrt{8})^m.$$

Now solve for x:

$$2\sqrt{3}x = (1 + \sqrt{3})(2 + \sqrt{3})^n - (1 - \sqrt{3})(2 - \sqrt{3})^n$$

and one of

$$2\sqrt{8}x = \begin{cases} (1 + \sqrt{8})(3 + \sqrt{8})^m - (1 - \sqrt{8})(3 - \sqrt{8})^m \\ (-1 + \sqrt{8})(3 + \sqrt{8})^m - (-1 - \sqrt{8})(3 - \sqrt{8})^m \end{cases}$$

But these two options are actually identically the same thing! So we get that $\sqrt{8}(1 + \sqrt{3})(2 + \sqrt{3})^n$ is exponentially close to $\sqrt{3}(\pm 1 + \sqrt{8})(3 + \sqrt{8})^m$. This is a linear form in logarithms! There exists m, n so that

$$m \log(3 + \sqrt{8}) - n \log(2 + \sqrt{3}) + \beta$$

is exponentially small, contradicting inhomogeneous Baker. Actually computing the coefficients, one finds that $m \leq 10^{487}$,then you have to check all the cases up to that point. But Baker and Davenport check a lot of these all at once with great efficiency, using continued fractions cleverly.

Now, let's generalize this result in the following simple way. Let E be the elliptic curve defined by $y^2 = (x - a)(x - b)(x - c)$, $a, b, c \in \mathbb{Z}$. Then we try to solve the system

$$x - a = A\square$$
$$x - b = B\square$$
$$x - c = C\square.$$

So we do the same trick as before, and we find that there are finitely many solutions and that they are effectively computable. The general case is where the elliptic curve doesn't have full 2-torsion over $\mathbb{Q}$, but we'll not do that here.

Application 3: The unit equation in a number field. Let K be a number field of degree $n = r + 2s$, r real embeddings and $2s$ complex embeddings. The unit group is then of rank $t = r + s - 1$.

Question 2 (unit equation)**.** *Find all solutions to $u + v = 1$, where u, v are units in K.*

Theorem 15. *There are finitely many solutions to the unit equation, and they can be effectively determined.*

Proof. Assume $\eta_1, \ldots, \eta_t$ are the fundamental units in K. Then any unit is of the form $u = \zeta \eta_1^{a_1} \cdots \eta_t^{a_t}$, where ζ is a root of unity, and the $a_j \in \mathbb{Z}$. We define the regulator $R = \det(\log |\eta_i^j|) \neq 0$. Let $\theta \in K$. Since there are $r + 2s$ embeddings of K, we have

$$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$$

real embeddings, and

$$\theta^{(r+1)}, \theta^{(r+2)}, \ldots, \theta^{(r+s)}$$

complex embeddings, and the

$$\theta^{(r+s+1)}, \theta^{(r+s+2)}, \ldots, \theta^{(r+2s)}$$

complex conjugates thereof. So if we know the first $r + s - 1$ of these we can determine all of them because we know the norm of θ. The regulator is a $t \times t$ determinant, so the definition of regulator makes sense.

Let $v = \zeta' \eta_1^{b_1} \cdots \eta_t^{b_t}$. Let's forget about the roots of unity for a minute here, they're not really the point of the proof here. So we know

$$\eta_1^{a_1} \cdots \eta_t^{a_t} + \eta_1^{b_1} \cdots \eta_t^{b_t} = 1.$$

We have a similar equation for each embedding. We know for each $1 \leq j \leq t$

$$\eta_1^{(j)a_1} \cdots \eta_t^{(j)a_t} + \eta_1^{(j)b_1} \cdots \eta_t^{(j)b_t} = 1.$$

We want to bound $||a_i||_{l^2}$ or $||a_i||_{l^2}$, so we want two big real numbers adding to zero, so look at the regulator.

$$\begin{pmatrix} \log|\eta_1^{(1)}| & \cdots & \log|\eta_t^{(1)}| \\ \log|\eta_1^{(2)}| & \cdots & \log|\eta_t^{(2)}| \\ \vdots & & \vdots \\ \log|\eta_1^{(t)}| & \cdots & \log|\eta_t^{(t)}| \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ a_t \end{pmatrix} = (\log|\eta_1^{(1)a_1}| \cdots |\eta_t^{(1)a_t}|, \ldots, \log|\eta_1^{(t)a_1} \cdots \eta_t^{(t)a_t}|).$$

So we have that

$$||\log|\eta_1^{(i)a_1}| \cdots |\eta_t^{(i)a_t}|||_{l^2} \gg R||a_i||_{l^2}.$$

We want to be able to say that there exists a j so that $\eta_1^{(j)a_1} \cdot \eta_t^{(j)a_t}$ is big in absolute value. And by the above, we get this. Why do we need u, v large? Because $\log u = \log(1-v) \approx \log v$ makes sense if v is large, then we can derive a contradiction using effective Baker's theorem.

□

Application 4: the Thue equation. Let K be a number field, and $\alpha_1, \ldots, \alpha_d \in \mathcal{O}_K$ be distinct, $d \geq 3$. Fix $\mu \in \mathcal{O}_K$. Solve $(x - \alpha_1 y) \cdots (x - \alpha_d y) = \mu$, with $x, y \in \mathcal{O}_K$. For example, $x^3 - 2y^3 = 73$.

Theorem 16. *The Thue equation (above) has finitely many solutions, and can be effectively solved.*

We can reduce powers in the Thue equation $\pmod p$ and get an equation of the form $\alpha\xi^p + \beta\nu^p = 1$, with finitely many choices for α, β. This looks just like the unit equation. So the Thue and unit equations are closely related.

10 The Thue Equation, Hyperelliptic Equations

Last time we discussed the unit equation. Let K be a number field of degree n. The equation $u + v = 1$ with units u and v has finitely many solutions, and they can be effectively determined. We can ask more generally for solutions to $\alpha u + \beta v = \gamma$ with $\alpha, \beta, \gamma \in \mathcal{O}_K$. By the same method, there are finitely many and they can be effectively determined. Zimmeit showed that there is a universal constant for regulators, something like $R \geq 0.056...$ universally!

The next application is the Thue equation. Again, let K be a number field of degree $n = r + 2s$, and $t = r + s - 1$ which is the rank of the unit group. Let $\alpha_1, \ldots, \alpha_d$ be distinct algebraic integers, $d \geq 3$. We want to find solutions to

$$(x - \alpha_1 y) \cdots (x - \alpha_d y) = \mu$$

with $0 \neq \mu \in \mathcal{O}_K$. We want to show that this has only finitely many solutions for $x, y \in \mathcal{O}_K$ which can be effectively determined (i.e. there is a bound for these solutions). e.g. the equation $x^3 - 2y^3 = k$.

Corollary 7. *If α is algebraic of degree n, then there is some function $g(q) \to \infty$ as $q \to \infty$ such that $|\alpha - \frac{a}{q}| \geq \frac{g(q)}{q^n}$, where $g(q)$ is explicitly effectively computable.*

Baker: $g(q) = \exp((\log q)^\delta)$ for some δ. Feldman: $g(q) = q^\delta$. So the proof of the corollary is to apply the Thue theorem to $|(\alpha q - a)(\alpha_2 q - a) \cdots (\alpha_n q - a)| \to \infty$ as $q \to \infty$. So the Thue theorem shows that the left hand side is at least as large as something $\to \infty / q^n$.

Now we try to prove Thue's theorem. A crucial fact in the proof is that $N(x - \alpha_i y) | N(\mu)$. There are many ways of defining height of an algebraic number. Here's a new one we will use:

$$\overline{|\alpha|} = \max |\alpha^{(j)}|,$$

where the $\alpha^{(j)}$ are the Galois conjugates of α. We have

Theorem 17 (Northcott)**.** *There are only finitely many algebraic integers α of bounded degree and bounded $\overline{|\alpha|}$.*

Proof. Write down the polynomial $(x - \alpha^{(1)}) \cdots (x - \alpha^{(n)})$, and observe that it has $\mathbb{Q}$ coefficients which are bounded. There are only finitely many monic polynomials in integers with bounded coefficients and bounded degree, which therefore have a finite number of solutions. □

We now proceed to the proof of the Thue equation. Suppose we have a solution $(x - \alpha_j y) = \beta_j u_j$, with $u_j \in \mathcal{O}_K^\times$ and $\overline{|\beta_j|}$ bounded. As in last lecture, let us denote the Galois conjugates of an algebraic number $\theta \in K$ as

$$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$$

real embeddings, and

$$\theta^{(r+1)}, \theta^{(r+2)}, \ldots, \theta^{(r+s)}$$

complex embeddings, and the

$$\theta^{(r+s+1)}, \theta^{(r+s+2)}, \ldots, \theta^{(r+2s)}$$

complex conjugates thereof. We can then rig the u_j such that $(x - \alpha_j y)^{(k)}$ is bounded for $k = 1, \ldots, r + s - 1$. But we also have $N(x - \alpha_j y) | N(\mu)$ so that $(x - \alpha_j y)^{(k+s)}$ is bounded also, and hence all the conjugates. Then we have

$$\begin{aligned} x - \alpha_1 y &= \beta_1 u_1 \\ x - \alpha_2 y &= \beta_2 u_2 \\ &\cdots \\ x - \alpha_d y &= \beta_d u_d \end{aligned}$$

with $\beta_1, \ldots, \beta_d$ fixed of bounded height. We want to show that the u_j are determined. We can take linear combinations of the above, and get, for example with the first three

$$\begin{aligned} \alpha_2 y - \alpha_1 y &= \beta_1 u_1 - \beta_2 u_2 \\ \alpha_3 y - \alpha_1 y &= \beta_1 u_1 - \beta_3 u_3 \end{aligned}$$

and solving for y between these two get

$$(\alpha_3 - \alpha_1)(\beta_1 u_1 - \beta_2 u_2) = (\alpha_2 - \alpha_1)(\beta_1 u_1 - \beta_3 u_3).$$

This is a unit equation in the variables $u_2/u_1, u_3/u_1$, which we already know has finitely many solutions, effectively determined. But choosing $1, 2, 3$ was arbitrary here, so we actually know that each of u_i/u_1 have effectively finitely many solutions. But we have

$$\mu = \beta_1 \cdots \beta_d u_1^d \left(\frac{u_2}{u_1}\right) \cdots \left(\frac{u_d}{u_1}\right),$$

so u_1 is determined also! So applying Northcott's theorem, that solves the Thue equation.

Application 5: Integer solutions to elliptic and hyperelliptic curves. Let K be a number field and $\alpha_1, \ldots, \alpha_d$ be $d \geq 3$ distinct algebraic integers. The problem is to effectively solve

$$y^2 = (x - \alpha_1) \cdots (x - \alpha_d)$$

effectively. (N.B. This gives an excellent result on the size of the integer solutions to such an equation, but no information as to the number of solutions. That is an entirely different problem to be attacked by entirely different means. But this goes to show how many questions one can ask.) Our approach is similar to descent, we try to write each of these as something fixed times a square. Actually, we will work with ideals. As ideals, we have

$$(y^2) = (x - \alpha_1) \cdots (x - \alpha_d),$$

and

$$(x - \alpha_j) = \mathfrak{a}_j \mathfrak{b}_j^2,$$

and the only things that could lie in $\mathfrak{a}_j$ come from coprimality conditions. So $\mathfrak{a}_j$ divides $\prod_{k \neq j} (\alpha_j - \alpha_k)$, so there can only be finitely many choices for $\mathfrak{a}_j$. Let $\mathfrak{a}_j^{-1}$ and $\mathfrak{b}_j^{-1}$ be integral ideals inverse to $\mathfrak{a}_j$ and $\mathfrak{b}_j$ in the ideal class group with bounded norm. So we have

$$\mathfrak{a}_j^{-1} \mathfrak{b}_j^{-2} (x - \alpha_j) = (\mathfrak{a}_j \mathfrak{a}_j^{-1})(\mathfrak{b}_j \mathfrak{b}_j^{-1})^2$$

The ideals on the right hand side are principal, and all of the $\mathfrak{a}_j, \mathfrak{a}_j^{-1}, \mathfrak{b}_j, \mathfrak{b}_j^{-1}$ have bounded norm. Then as elements, we have

$$x - \alpha_j = \frac{A_j}{B_j} \gamma_j^2 \epsilon_j,$$

where A_j, B_j have bounded height in the sense of Thue, γ_j is an algebraic integer, and ϵ_j is a unit. The B_j comes from $\mathfrak{a}_j^{-1} \mathfrak{b}_j^{-2}$, the A_j comes from $\mathfrak{a}_j \mathfrak{a}_j^{-1}$), and the γ^2 comes from $(\mathfrak{b}_j \mathfrak{b}_j^{-1})^2$. But we still have to get rid of the unit. We have

$$\epsilon_j = \zeta \eta_1^{a_1} \cdots \eta_t^{a_t},$$

say. Take the squarefree part of this by absorbing the squares into γ_j^2. The conclusion is then that

$$x - \alpha_j = \frac{C_j}{D_j} \beta_j^2,$$

with $\overline{|C_j|}, \overline{|D_j|}$ bounded. Now, let's look at the first few of these equations we get and try to sort out what happens. We have

$$x - \alpha_1 = \frac{C_1}{D_1} \beta_1^2$$

$$x - \alpha_2 = \frac{C_2}{D_2}\beta_2^2$$
$$x - \alpha_3 = \frac{C_3}{D_3}\beta_3^2$$

so that

$$\alpha_2 - \alpha_1 = \frac{C_1}{D_1}\beta_1^2 - \frac{C_2}{D_2}\beta_2^2$$
$$\alpha_3 - \alpha_1 = \frac{C_1}{D_1}\beta_1^2 - \frac{C_3}{D_3}\beta_3^2$$
$$\alpha_3 - \alpha_2 = \frac{C_2}{D_2}\beta_2^2 - \frac{C_3}{D_3}\beta_3^2.$$

Now we're back in the situation of Baker-Davenport: simultaneous pell-type equations. So we use the same method to prove that the system has only finitely many solutions. First, clear denominators:

$$D_1D_2D_3(\alpha_2 - \alpha_1) = D_3D_2C_1\beta_1^2 - D_3D_1C_2\beta_2^2$$
$$D_1D_2D_3(\alpha_3 - \alpha_1) = D_3D_2C_1\beta_1^2 - D_1D_2C_3\beta_3^2$$
$$D_1D_2D_3(\alpha_3 - \alpha_2) = D_3D_1C_2\beta_2^2 - D_1D_2C_3\beta_3^2.$$

Now, factor. The right hand sides equal

$$(\sqrt{C_1D_2D_3}\beta_1+\sqrt{C_2D_1D_3}\beta_2)(\sqrt{C_1D_2D_3}\beta_1-\sqrt{C_2D_1D_3}\beta_2) := (G_{12}v_{12})(F_{12}u_{12})$$
$$(\sqrt{C_1D_2D_3}\beta_1+\sqrt{C_3D_2D_3}\beta_3)(\sqrt{C_1D_2D_3}\beta_1-\sqrt{C_3D_2D_3}\beta_3) := (G_{13}v_{13})(F_{13}u_{13})$$
$$(\sqrt{C_2D_1D_3}\beta_2+\sqrt{C_3D_2D_3}\beta_3)(\sqrt{C_2D_1D_3}\beta_2-\sqrt{C_3D_2D_3}\beta_3) := (G_{23}v_{23})(F_{23}u_{23}).$$

Let's now work in an extension of K which includes these square roots. Observe that the sum of these three equations is zero (left hand side). Now, multiply the three equations together. The right hand side is a product of 6 factor, each of which has effectively bounded norm. So, we may write each as a unit times a number which is effectively determined (hence the meaning of the above). i.e. The $\overline{|F_{12}|}, \overline{|F_{13}|}, \overline{|F_{23}|}, \overline{|G_{12}|}, \overline{|G_{13}|}, \overline{|G_{23}|}$ are bounded. Adding them, we get zero, so

$$F_{12}u_{12} - F_{13}u_{13} + F_{23}u_{23} = 0,$$

also

$$G_{12}v_{12} - G_{13}v_{13} = F_{23}u_{23},$$

etc. So given u_{23}, u_{13} and u_{12} are also determined. So this fixes what v_{23} is, and hence each of the v_{ij}. Therefore, we find that effectively finding integer solutions to hyperelliptic equations reduces to the Thue equation, and therefore we are finished.

So what have we got? If we are given an equation $y^2 = x^3 + ax + b$ with $|a|, |b| \leq H$, we have that the integer solutions have $\max(|x|, |y|) \leq \exp((10^6H)^{10^6})$, in general. In specific cases we can do much better, for example, the famous Mordell equation $y^2 = x^3 + k$ we have $\max(|x|, |y|) \leq \exp(ck^{1+\epsilon})$. There is also a famous problem

Conjecture 2 (Hall)**.** *For all coprime integers x, y, we have*

$$|y^2 - x^3| \gg x^{1/2-\epsilon}.$$

Baker's theorem implies a lower bound of $\gg (\log x)^{1-\epsilon}$, effectively. Also, the ABC conjecture implies Hall's conjecture. We'll talk more about this next class.

There is also a p-adic version of all of Baker's theorem. We want to say that a linear form in logarithms can't be too small in the p-adic metric. To treat this case, we need to take a large space. Start with $\mathbb{Q}_p$, then take the algebraic closure $\overline{\mathbb{Q}_p}$, which is no longer complete, so complete it again to get Ω_p. There is a theory of analytic functions on this object, developed by Mahler, Schnirelman, Sprindzuk, Brumer, Coates, Vanderpoorten, and Kun-Rui Yu. There's a nice book on applications of Baker's theorem and the p-adic Baker theorem by Shorey and Tijdeman called "Exponential Diophantine Equations". We won't go into any detail, so you should look there if you're interested.

To get a p-adic version of Baker's theorem, we'll have to get some analogue of the Cauchy integral formula or the maximum principle. Suppose

$$f(z) = \sum_{k=0}^{\infty} a_k (z-a)^k$$

is an analytic function converging for some $|z-a|_p < \rho \in \mathbb{R}$. If $(n,p) = 1$, then we factor

$$x^n - 1 = \prod_{i=1}^{n} (x - \zeta_i)$$

in Ω_p. Then we have some sort of circle, so we can write down an analogue for the Cauchy integral formula. Consider the limit

$$\lim_{\substack{n\to\infty \\ (n,p)=1}} \frac{1}{n} \sum_{j=1}^{n} f(a + r\zeta_j)$$

with $0 \neq r \in \Omega_p$. We call this limit

$$\int_{a,r} f(z)\, dz,$$

and you should think of this as similar to

$$\frac{1}{2\pi i} \int_{|z-a|=r} f(z) \frac{dz}{z-a}.$$

This integral has nice properties. For example, if $f(z)$ is analytic, then the integral evaluates to $f(a)$. There is also a maximum modulus principle, but it's now obvious by the ultrametric triangle inequality:

$$\left| \int_{a,r} f(z)\, dz \right|_p \leq \max_{|z|_p = |r|_p} |f(a+z)|_p.$$

Now, the rest of the proof of Baker's theorem goes through line for line. The statement at the end becomes

Theorem 18 (p-adic Baker's theorem)**.** *Let K a degree d number field, and let $\alpha_1, \ldots, \alpha_n \in K$ be nonzero, with $\overline{|\alpha_1|} \leq A_1, \ldots, \overline{|\alpha_n|} \leq A_n$. Let $\mathfrak{p}$ be a prime above p, and $b_1, \ldots, b_n \in \mathbb{Z}$ with $|b_j| \leq B$. Then*

$$ord_{\mathfrak{p}}(\alpha_1^{b_1} \cdots \alpha_n^{b_n} - 1) \leq (Cnd)^{Cn} \frac{p^d}{\log p} (\log A_1) \cdots (\log A_n)(\log B)^2$$

For comparison, Baker's theorem would say

$$|\alpha_1^{b_1} \cdots \alpha_n^{b_n} - 1| \gg \exp(-C(\log A_1) \cdots (\log A_n)(\log B)),$$

without the square on the last factor.

11 Effective p-adic Baker and Applications

Last time we stated a version of Baker in the p-adic case, but gave no proofs whatsoever. Here's the statement again:

Theorem 19 (Baker's theorem for p-adic valuations)**.** *Let K be a number field of degree d over $\mathbb{Q}$. Let $\mathfrak{p}$ be a prime of K over p. Let $\alpha_1, \ldots, \alpha_n \in \mathcal{O}_K$ of heights $\leq A_1, \ldots, A_n$ respectively. Let $b_1, \ldots, b_n \in \mathbb{Z}$, all $\leq B$. Then*

$$ord_{\mathfrak{p}}(\alpha_1^{b_1} \cdots \alpha_n^{b_n} - 1) \leq (Cnd)^{cn} \frac{p}{\log p} (\log A_1) \cdots (\log A_n)(\log B)^2.$$

This theorem is due to the combined work of many people: Yu, Vanderpoorten, etc. It is slightly weaker than the corresponding result at the archimedian place due to the $(\log B)^2$ appearing above instead of $\log B$.

This result is very useful, as it allows us to solve the S-unit equation, $u+v = 1$ where u and v are S-units. S is a fixed finite set of primes. In the archimedian case, we would interpret S as the set of all infinite places. We get that the S-unit equation has only finitely many solutions, and that they can be effectively determined. What a nice theorem!

For example suppose we have the following problem: Find all solutions $(x, y, z) \in \mathbb{N}^3$, with $x + y = z$ satisfying the statement

$$p|xyz \implies p \in S.$$

We can do this with our result on the S-unit equation, and get an effective upper bound on the size of the solutions.

A Here's a different problem: Count the number of solutions to the above equation. A result of Evertse says that the number of solutions is $\ll \exp(a|S| + b)$, but his result is ineffective. Thus, we can tell that there are finitely many solutions, but we don't know how many. We've stated Evertse's result over $\mathbb{Q}$,

but it works with different constants for any other number field. Evertse only depends on the rank of the group.

Now, we do another application. First, we take the Thue equation:

$$(x - \alpha_1 y) \cdots (x - \alpha_d y) = \mu,$$

where $d \geq 3$ and we solve in algebraic integers. Now, we will replace μ and get the Thue-Mahler equation. Take μ to be comprised of primes in a fixed finite set S. That is, we take μ to be varying instead of fixed, as it is in the Thue equation. Said differently: the left hand side is allowed to be any S-integer. This is related to the S-unit equation $x + y = z$. If we let $x = aA^4$, and $y = bB^4$, then $aA^4 + bB^4 = z$. So if we factor the the left, we have $4^{|S|}$ from factoring a, b, and if we forget A and B are composed of primes in S at the end we get an equation which looks like Thue-Mahler. Overkill: $aA^4 + bB^4 = cC^4$ is a curve of genus ≥ 2 so Falting's theorem implies there are only finitely many solutions.

Conjecture 3 (Erdös, Stewart, Tijdeman)**.**

$$x + y = z,\ p|xyz \Longrightarrow p \in S$$

has at most $\exp(|S|^{2/3+o(1)})$ *solutions as* $|S| \to \infty$*.*

Example 1 (Konyagan and Soundararajan)**.** *There exists* S *with* $\geq \exp(|S|^{2-\sqrt{2}-\varepsilon})$ *solutions.*

This is in a similar vein to an old result of Erdös, Stewart and Tijdeman which says that there exists a special set of primes S with $\geq \exp(|S|^{1/2})$ solutions. Furthermore, Jeff Lagarias and Sound have a result that is $\mathcal{S}$ is the set of the first $|\mathcal{S}|$ primes, then there are $\geq \exp(|\mathcal{S}|^{1/8})$ solutions.

Here's another application. We have

Conjecture 4 (ABC conjecture)**.** *Given* $a, b, c \in \mathbb{N}$ *with* $a+b = c$ *and* $(a, b) = 1$*, then*

$$\max(|a|, |b|, |c|) \leq c_\varepsilon \left(\prod_{p|abc} p \right)^{1+\varepsilon}.$$

This conjecture is already quite deep because it would imply Fermat's last theorem, and all sorts of other things. We can get some sort of result along these lines, however, by bounding the right hand side of the above from below using the effective p-adic Baker. There is a a result

Theorem 20 (Stewart-Tijdeman)**.** *Given the same assumptions as the ABC conjecture we have,*

$$\max(|a|, |b|, |c|) \ll \exp\left(\left(\prod_{p|abc} p \right)^{15} \right).$$

The bound is pretty bad, but at least we have something. There is a result of Yu and Stewart which gets the constant down from 15 to 1/3, but it's proof is considerably more involved.

Proof. Suppose a, b, c is composed of the primes $p_1 \leq \ldots \leq p_r$. Let $a = p_1^{a_1} \cdots p_r^{a_r}$, $b = p_1^{b_1} \cdots p_r^{b_r}$, and $c = p_1^{c_1} \cdots p_r^{c_r}$. We want a lower bound for $R := \prod_{j=1}^{r} p_j$. By writing $a + b = c$ as $\frac{b}{c} = 1 - \frac{a}{c}$, and similarly for $\frac{a}{c}$, we can use p-adic Baker to get

$$B := \max(a_1, \ldots, a_r, b_1, \ldots, b_r, c_1, \ldots, c_r) \leq (Cr)^{Cr} \frac{p_r}{\log p_r} (\log p_1) \cdots (\log p_r)(\log B)^2.$$

But $R \geq$ the product of the first r primes, which is $\approx \exp(r \log r)$ by the prime number theorem. So $(Cr)^{Cr}$ is bounded by a power of R, say by R^A. We also have $\frac{p_r}{\log p_r} \leq R$, and

$$(\log p_1) \cdots (\log p_r) \leq p_1 \cdots p_r \leq R.$$

So, $B \leq R^c$ for some constant c. So $abc \leq R^{R^c} \leq \exp(R^{c+1})$. □

A few miscellaneous remarks: First, we show that there are sets of primes S so that the S-unit equation has as many as $\exp(\sqrt{|S|})$ solutions.

Let's construct an example to show this. Let $\mathcal{S}(y)$ be the set of y-smooth numbers. Recall that a number n is y-smooth if $p|n \implies p \leq y$. Let

$$\psi(x, y) = \sum_{\substack{n \leq x \\ n \in \mathcal{S}(y)}} 1$$

Consider y-smooth numbers $a, b \leq X$. There are $\psi(X, y)$ of these (ignoring coprimality conditions). The sums $a + b$ run up to $2X$. So there exists a "popular" c with at least $\frac{\psi(X,y)^2}{2X}$ representations as $a + b = c$. Let $S = \{p \leq y\} \cup \{p|c\} \sim \frac{y}{\log y} + \frac{\log x}{\log \log x}$, i.e. the primes dividing each of a, b, c. Then if we take $y = (\log x)^\alpha$, we have $\psi(x, y) = x^{1 - \frac{1}{\alpha} + o(1)}$ so long as $\alpha > 1$. Take $\alpha \geq 2 + \varepsilon$. Then the popular c has $\geq x^\varepsilon = \exp(y^{1/2 - \varepsilon})$ representations.

The next miscellaneous remark is that we need the ε in the ABC conjecture, otherwise it is false.

Now, consider the y-smooth numbers up to x, $\psi(x, y)$ of them. The idea here is that there are two which are very close, within $x/\psi(x, y)$ of each other (but we'll do something slightly more refined. We choose a and c to be y-smooth, and close together, so that $b = c - a$ is smallest. To avoid coprimality conditions, let's look at the interval $[(1 + \delta)^j, (1 + \delta)^{j+1}]$ in lieu of the interval $[1, x]$ to make sure the gcd works out. The number of such intervals is $\log x/\delta$. If $\psi(x, y) > \frac{\log x}{\delta}$, we can find $(a, c) = 1$ with $a, c \in [(1 + \delta)^j, (1 + \delta)^{j+1}]$. Now, $b = c - a$, so $\max(a, b, c) = c$ and

$$R := \prod_{p|abc} p \leq \left(\prod_{p \leq y} p \right) b \leq \delta c e^y.$$

The e^y comes from the product. We want R small. We pick δ so that there are at least two points in one interval, so maybe up to a constant, $R \lesssim \frac{c \log x e^y}{\psi(x,y)}$. So we use calculus to minimize this. To do this, we first have to find a lower bound for $\psi(x,y)$. So, we consider all y-smooth numbers, that is, consider all possible choices of $k_p \in \mathbb{N}$ for each $p \leq y$, and such that $\prod_{p \leq y} p \leq x$. We want to count the number of possible choices for $\{k_p\}$ subject to these conditions. So, we have

$$\sum_{p \leq y} k_p \log p \leq \log x$$

so

$$\sum_{p \leq y} k_p \leq \frac{\log x}{\log y}.$$

Thus we've reduced the problem of counting y-smooth numbers to the problem to counting lattice points in a high dimensional tetrahedron. More precisely,

$$\psi(x,y) \geq \#\{(k_p)_{p \leq y} | k_p \in \mathbb{N}, \sum_{p \leq y} k_p \leq \frac{\log x}{\log y}\}.$$

We can count the lattice points by appealing to simple estimates about the volume of a high-dimensional tetrahedron. So we get the lower bound

$$\psi(x,y) \geq \frac{\left(\frac{\log x}{\log y}\right) \pi(y)}{\pi(y)!} = \left(\frac{e \log x}{y}\right)^{\frac{y}{\log y}}.$$

Because e^y is $\approx y^{y/\log y}$

$$R \leq \frac{c \log x e^y}{\psi(x,y)} \leq c \log x \left(\frac{y^2}{e \log x}\right)^{y/\log y},$$

choosing $y = \sqrt{\log x}$, this is

$$R \leq c(\log x) \exp\left(\frac{-2\sqrt{\log x}}{\log \log x}\right),$$

where c is the max of a, b, c. This bound gives a counterexample to the ABC conjecture if we remove the ε.

Baker and Granville: quantitative version of ABC

Mason: ABC for polynomials

Cartan: ABC for holomorphic functions

We'll discuss two more results before going on to diophantine approximation, the second half of the course.

1. Six exponentials theorem
2. The Schneider-Lang theorem

A putnam problem: If $1^\alpha, 2^\alpha, 3^\alpha, \ldots \in \mathbb{N}$, prove that $\alpha \in \mathbb{N}$.

A corollary of the six exponentials theorem is that if $\alpha \notin \mathbb{Q}$ then one of $2^\alpha, 3^\alpha$, and 5^α is transcendental.

Conjecture 5. *In fact, one of 2^α and 3^α is transcendental.*

Theorem 21 (old, first published accounts due to Ramachandra and Lang)**.** *Let $\alpha_1, \alpha_2 \in \mathbb{C}$, linearly independent over $\mathbb{Q}$. And $\beta_1, \beta_2, \beta_3 \in \mathbb{C}$, also linearly independent over $\mathbb{Q}$. Then one of $\exp(\alpha_i \beta_j)$ is transcendental.*

Conjecture 6. *Instead consider only $\beta_1, \beta_2 \in \mathbb{C}$, linearly independent over $\mathbb{Q}$. Then one of the four $\exp(\alpha_i \beta_j)$ are transcendental.*

proof (corollary of 6 exponentials theorem). Let $\beta_1 = \log 2$, $\beta_2 = \log 3$, and $\beta_3 = \log 5$. Let $\alpha_1 = 1$, $\alpha_2 = \alpha$. □

Corollary 8. *If β is transcendental, there are at most 2 algebraic numbers, multiplicatively independent, for which α_1^β and α_2^β are algebraic.*

The corollary complements the Gelfond-Schneider theorem.

12 Six Exponentials Theorem

Six exponentials theorem: If α_1, α_2 are linearly independent over $\mathbb{Q}$, and $\beta_1, \beta_2, \beta_3$ are linearly independent over $\mathbb{Q}$, then one of the six $\exp(\alpha_i \beta_j)$ is transcendental. A corollary is that one of $2^\alpha, 3^\alpha, 5^\alpha$ is transcendental for $\alpha \notin \mathbb{Q}$. e.g. one of $2^\pi, 2^{\pi^2}, 2^{\pi^3}$ is transcendental.

Proof. We construct an auxiliary function

$$\phi(z) = \sum_{k_1, k_2 = 0}^{K} p(k_1, k_2) e^{k_1 \alpha_1 z} e^{k_2 \alpha_2 z}.$$

We want to pick the $p(k_1, k_2)$ to be not all zero, lie in $\mathcal{O}_F$, and $\overline{|p(k_1, k_2)|}$ small. So we want $l_1\beta_1 + l_2\beta_2 + l_3\beta_3$, for $1 \leq l_1, l_2, l_3 \leq L$ to have $\phi(l_1\beta_1 + l_2\beta_2 + l_3\beta_3) = 0$. We'll see that this is easier to do that it was to prove Baker's theorem, as we have L^3 equations, and K^2 free variables, so we'll eventually take $K^2 \geq 2L^3$. We use the Thue-Siegel lemma again. Actually we need a slight modification of the Thue-Siegel lemma for a number field.

Lemma 4 (Thue-Siegel for a number field)**.** *Let F be a number field, and consider M variables. Suppose we have a homogeneous linear equation*

$$\sum_{j=1}^{N} a_{ij} x_j = 0,$$

with $N > M$ and $\alpha_{ij} \in \mathcal{O}_F$. Assume that $\overline{|a_{ij}|} \leq A$. Then there exists a nontrivial solution with

$$\overline{|x_j|} \leq (CNA)^{\frac{M}{N-M}},$$

where the constants only depend on F and nothing else.

Proof. The proof is the same as in the rational case. Let $w_1, \ldots, w_d$ be an integral basis for $\mathcal{O}_F$. Write a_{ij} and x_j in terms of $w_1, \ldots, w_d$. Then we have Md equations and Nd variables, and the size of the w_i are fixed in terms of F, so we can bound them by CA. Now, apply the Thue-Siegel lemma. □

Now recall we were in the middle of constructing

$$\phi(z) = \sum_{k_1,k_2=0}^{K} p(k_1,k_2) e^{k_1\alpha_1 z} e^{k_2\alpha_2 z}.$$

Evaluating ϕ we will get powers of $e^{\alpha_i\beta_j}$ going up to KL. To produce a contradiction, assume all $e^{\alpha_i\beta_j}$ are algebraic. Clear denominators by multiplying through by, say, D^{6KL}. So $D^{6KL}\phi(z)$ vanishes at the L^3 points $l_1\beta_1+l_2\beta_2+l_3\beta_3$. The size of the coefficients is C^{KL}, so by the Thue-Siegel lemma for number fields, we can find $p(k_1,k_2)$ with

$$\overline{|p(k_1,k_2)|} \leq (C^{KL})^{\frac{L^3}{K^2-L^3}}.$$

Let $K^2 = 2L^3$, then $\overline{|p(k_1,k_2)|} \leq C^{KL}$. Fact: ϕ is not identically zero, because α_1, α_2 are linearly independent over $\mathbb{Q}$. Fact: ϕ does not vanish on all linear combinations $l_1\beta_1 + l_2\beta_2 + l_3\beta_3$ with $l_1, l_2, l_3 \in \mathbb{N}$. Why? $\phi(z)$ is holomorphic, and these points are dense in $\mathbb{C}$. Alternately, because $\phi(z)$ is order 1, and can only have about R zeros in a circle of radius R, but it has at least R^3 zeros. So there is a number $s \geq L$ such that ϕ vanishes at all $l_1\beta_1 + l_2\beta_2 + l_3\beta_3$ with $l_j < s$ but doesn't vanish for some chosen $W = s_1\beta_1 + s_2\beta_2 + s_3\beta_3$, with $\max(s_1, s_2, s_3) = s$.

Now look at

$$\frac{\phi(z)}{\prod_{l_1,l_2,l_3<s}(z - l_1\beta_1 - l_2\beta_2 - l_3\beta_3)};$$

let $z = s_1\beta_1 + s_2\beta_2 + s_3\beta_3$, and use maximum modulus principle on some circle $|z| = R$. Then we have

$$|\phi(s_1\beta_1 + s_2\beta_2 + s_3\beta_3)| \leq (Cs)^{s^3} \max_{|z|=R} \frac{\phi(z)}{\prod_{l_1,l_2,l_3<s}(z - l_1\beta_1 - l_2\beta_2 - l_3\beta_3)}$$

$$\leq \frac{(Cs)^{s^3}}{(R/2)^{s^3}} \max_{|z|=R} |\phi(z)| \leq \frac{(Cs)^{s^3}}{(R/2)^{s^3}} C^{KL} \exp(CRK).$$

Choose $R = s^3/K$. Then the above is

$$\leq C^{KL} \left(\frac{10CK}{s^2}\right)^{s^3} \leq \exp(-cs^3 \log s),$$

where we've used $s > L$, $K = 2^{1/2}L^{3/2}$. So if all of it's conjugates are not too big, the usual norm argument will show that it is actually zero. So let's do it! After multiplying by D^{6KL}, $D^{6KL}\phi(s_1\beta_1 + s_2\beta_2 + s_3\beta_3)$ is an algebraic integer, and by our estimate on $|p(k_1, k_2)|$, we have that all it's conjugates are $\leq C^{KL}\exp(CKs)D^{6sK}$. So ϕ is zero, but not zero. Contradiction. $\square$

What about 4 exponentials? Then we'd have K^2 free variables, and L^2 equations. So we'd have to take $K = 2L$ in the end, and $s > L$ which would give $\left(\frac{(\cdots)K}{s}\right)^{s^2}$, and barely fail to give the 4 exponentials conjecture.

Another example from this circle of idea is the

Theorem 22 (Schneider-Lang)**.** *Let K be a number field, and $f_1, \ldots, f_N$ meromorphic functions of order $< \rho$. Let $f_i = g_i/h_i$, where g, h are holomorphic functions, and their orders are $< \rho$. Consider the ring $k[f_1, \ldots, f_N]$, and assume it satisfies two properties,*

1. *This ring has transcendence degree ≥ 2.*
2. *$\frac{d}{dz}$ preserves this ring. (N.B. surjectivity not required.)*

Then there are only finitely many $w_1, \ldots, w_m$ where the f_j are simultaneously algebraic. We have $m \leq 20\rho[K : \mathbb{Q}]$.

Before the proof, we do some applications of the Schneider-Lang theorem.

Example 1: Take $f_1 = z$, and $f_2 = e^z$. Then there are only finitely many $\alpha \in K$ with $e^\alpha \in K$. But if α has e^α algebraic, then $n\alpha$ is also algebraic for any $n \in \mathbb{N}$. So $e^\alpha \notin \overline{\mathbb{Q}}$ if $\alpha \neq= 0, \alpha \in \overline{\mathbb{Q}}$. So we recover a special case of Lindemann's theorem.

Example 2: Let $f_1 = e^z$, $f_2 = e^{\beta z}$, $\beta \in \overline{\mathbb{Q}}, \beta \notin \mathbb{Q}, \beta \in K$. The we get that there are only finitely many $\alpha \in K$ for which $\alpha^\beta \in K$ But if there is one, there are infinitely many: $\alpha, \alpha^2, \alpha^3, \ldots$, except if $\alpha = 0, 1$, i.e. we have recovered Gelfond-Schneider. (Unless α is a root of unity, but we can finesse this...)

Example 3: Let Λ be a lattice, say $\Lambda = \omega_1\mathbb{Z} + \omega_2\mathbb{Z}$, $\omega_2/\omega_1 \notin \mathbb{R}$. We have the doubly periodic function

$$\wp(z) = \frac{1}{z^2} + \sum_{0 \neq \lambda \in \Lambda} \left[\frac{1}{(z+\lambda)^2} - \frac{1}{\lambda^2}\right].$$

It is meromorphic, and has poles of order 2 at the points of Λ. Then

$$\wp'(z) = \frac{-2}{z^3} - \sum_{0 \neq \lambda \in \Lambda} \frac{2}{(z+\lambda)^3}$$

is also meromorphic of order two. We have the relation

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3,$$

where $g_2 = 60G_4 = 60\sum_{\lambda\neq=0}\frac{1}{\lambda^4}$, and $g_3 = 140G_6 = 140\sum_{\lambda\neq=0}\frac{1}{\lambda^6}$. Suppose we have a lattice with g_2 and g_3 algebraic, in some K. Then $K[\wp(z), \wp'(z), z]$ satisfies the conditions of the Schneider-Lang theorem. So there are only finitely many $\alpha \in K$ with $\wp(\alpha)$, $\wp'(\alpha)$ both in K. But using elliptic curve addition, we can add α's

Suppose we have periods ω_1, ω_2. Then consider $\wp(\omega_1/2)$ and $\wp'(\omega_1/2)$, and suppose that they are algebraic. (We would pick ω_1, but $\wp$ is not well defined there, so $\omega_1/2$ is the next best thing.) Siegel proved that at least one of the two periods are transcendental. Schneider proved both are. If $\omega_1/2, \wp(\omega_1/2)$ and $\wp'(\omega_1/2)$ are all algebraic, then $n\omega_1/2$ is also, contradicting Schneider-Lang. As a consequence, we know that if α is algebraic, then $\wp(\alpha)$ is transcendental.

Example 4: The modular j-function.

$$j(\tau) := 1728\frac{g_2^3}{g_2^3 - 27g_3^2},$$

where $\tau = \omega_2/\omega_1$, and thus the lattice is generated by 1 and τ. Consequence: if τ is algebraic and τ is not a quadratic irrationality, then $j(\tau)$ is transcendental. Note: If τ is a quadratic irrationality, then $j(\tau)$ generates the Hilbert class field of $\mathbb{Q}(\tau)$, so it has a very significant algebraic meaning!

Example 5: With much more work (due to Chudnovsky), one can show that the periods don't have any relation with π for E a CM elliptic curve. This in turn implies that $\Gamma(1/4), \Gamma(1/3), \Gamma(1/6)$ are transcendental by picking clever CM elliptic curves.

13 Schneider-Lang Theorem

Recall the Schneider-Lang theorem: Let $f_1, \ldots, f_N$ be meromorphic functions of order $< \rho$, K a number field such that

1. $K[f_1, \ldots f_N]$ has transcendence degree ≥ 2.
2. $K[f_1, \ldots f_N]$ is mapped into itself by $\frac{d}{dz}$.

If $\omega_1, \ldots, \omega_m$ are complex numbers not being the pole of any f_j and $f_j(\omega_j) \in K$ for all j, k. Then $m \leq 20\rho[K : \mathbb{Q}]$.

One corollary was the Gelfond-Schneider theorem after taking $f_1 = e^z$ and $f_2 = e^{\beta z}$, $\beta \notin \mathbb{Q}$, $\beta \in \overline{\mathbb{Q}}$. There was some confusion in this last time. If $n \log \alpha$, α algebraic then we get the Gelfond-Schneider theorem so long as $\alpha \neq 0$ and $\log \alpha \neq 0$.

Other examples: To every lattice $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$, with $\omega_2/\omega_1 \notin \mathbb{R}$ we look at doubly periodic functions on Λ. We have for the Weierstauss function (defined last time)

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3$$

where

$$g_2 := 60G_4 := 60\sum_{\lambda\neq 0}\frac{1}{\lambda^4}$$

and

$$g_3 := 140G_6 := 140 \sum_{\lambda \neq 0} \frac{1}{\lambda^6}.$$

Then if g_2, g_3 are algebraic, every period ω_1, ω_2 is transcendental (this is the result of Siegel / Schneider mentioned last time). Contrapositively, if α is algebraic, then $\wp(\alpha)$ is transcendental. This construction and a little more (if the elliptic curve is CM) shows that the periods are also independent of π.

Proof (of example 4). Take any lattice where $\omega_1/\omega_2 = \tau$. That is, take any lattice homothetic to this Λ. Suppose $j(\tau)$ is algebraic, and pick ω_1 and ω_2 on this lattice with g_2 and g_3 algebraic. So now we're in the previous situation. So for this lattice, we have

$$\wp'(z)^2 = 4\wp(z)^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda).$$

Let's work in a number field K which contains τ. Consider the values

$$\wp(z), \wp'(z), \wp(\tau z), \wp'(\tau z),$$

and plug in $z = (n+\frac{1}{2})\omega_1$. Then $(\wp(z), \wp'(z))$ are 2-torsion points, which means that they're algebraic. Now, $\tau z = (n+\frac{1}{2})\omega_2$ similarly gives algebraic values for $\wp(\tau z)$ and $\wp'(\tau z)$. So by the Schneider-Lang theorem $\wp(z)$ and $\wp(\tau z)$ are algebraically dependent. Now we can play around and match up the poles, so this forces $\tau \ell \omega_2 \in \Lambda$ for some integer ℓ (exercise). Now we're done, since $\frac{\ell\omega_2^2}{\omega_1} = a\omega_1 + b\omega_2$, so $\tau = \omega_2/\omega_1$ is imaginary quadratic. □

Proof (of Schneider-Lang). Out of $f_1, \ldots, f_N$ there are 2 functions which are algebraically independent, say f and g. Use these to construct an auxiliary function

$$\phi(z) = \sum_{k_1,k_2=1}^{K} p(k_1,k_2) f(z)^{k_1} g(z)^{k_2}.$$

The algebraic independence shows that this ϕ is no identically zero unless all $p(k_1,k_2)$ are zero. We will pick $p(k_1,k_2)$ to be algebraic integers in K with smallish size. Say $z = \omega_1, \ldots, \omega_m$ are points where $f_j(\omega_k) \in K$. We want that $\phi^{(\ell)}(\omega_j) = 0$ for all $0 \leq \ell \leq L$. By the second condition, for any j, f_j' is expressible as a polynomial in the other meromorphic functions, say, $f_j'(z) = P_j(f_1, \ldots, f_N)$. There are Lm equations to be satisfied, and K^2 free variables. What happens to the size of these quantities when we differentiate a bunch of times? Pick B large which kills all denominators of $f(\omega_j), g(\omega_j), f_j'(\omega_k), \ldots$ etc. We want $B^{2K+L}\phi^{(\ell)}(\omega_j) = 0$. The size of the coefficients is $\leq B^K(CK)^L$. So choose $K^2 = 2Lm$. The Thue-Siegel lemma apples, and we find $p(k_1,k_2)$ with $\overline{|p(k_1,k_2)|} \leq \exp(L \log L)$. Now use some version of the maximum modulus principle to get a contradiction in the usual way (how many times have done this now?).

Pick s to be the smallest number such that $\phi^{(s+1)}(\omega) \neq 0$ for some $\omega = \omega_1, \ldots, \omega_n$, but all smaller derivatives are zero. By construction, $s \geq L$. Look at

$$\frac{\phi(z)\Theta(z)^{2k}}{((z-\omega_1)\cdots(z-\omega_n))^{s+j}}$$

where $\Theta(z)$ is a holomorphic function of order $< \rho$ such that $f(z)\Theta(z)$ and $g(z)\Theta(z)$ are holomorphic. Then above fraction is an entire function of order $< \rho$. Apply the maximum modulus principle using a circle of big radius R to be chosen later. Evaluate at $z = w$. Then

$$\left|\frac{\phi^{(s+1)}(\omega)\Theta(\omega)^{2k}}{\prod_{\omega_j \neq \omega}(\omega-\omega_j)^{s+1}(s+1)!}\right| \leq \max_{|z|=R} \frac{\phi(z)\Theta(z)^{2k}}{((z-\omega_1)\cdots(z-\omega_n))^{s+1}} \leq \exp(CKR^\rho + L\log L - sm\log R/2),$$

where in the last inequality, the three terms come from Θ, ϕ and the denominator, respectively.

Recall we have $K^2 = 2Lm$ and $s \geq L$, so the optimal value of R is $C\rho K R^{\rho-1} = \frac{sm}{R}$. So $R = \left(\frac{sm}{K}\right)^{1/\rho}$. So the bound is

$$\leq \exp(L\log L - \frac{sm}{\rho}\log\frac{sm}{10K}).$$

Conclusion:

$$|\phi^{(s+1)}(\omega)| \leq \exp(2s\log s - \frac{sm}{\rho}\log\frac{sm}{10K}).$$

By multiplying $\phi^{(s+1)}(\omega)$ by a suitable B^{s+2K}, we get an algebraically integer which is $\leq \exp(L\log L + Cs)$, and we derive a contradiction by a norm calculation. The norm calculation implies $|\phi^{(s+1)}(\omega)| \geq \exp(-L\log L - dCs)$, where $d = [K:\mathbb{Q}]$. So if $m = 20\rho[K:\mathbb{Q}]$, we get the desired contradiction. Something like 4 probably still works in the place of 20. □

Next up: Diophantine approximation.

14 Introduction to Diophantine Approximation

Today we start Diophantine approximation, and the subject will take up the remainder of the course.

We have an algebraic number α of degree d. Assume it is real. Then we have

Theorem 23 (Dirichlet)**.** *There are infinitely many $p/q \in \mathbb{Q}$ with $|\alpha - p/q| \leq 1/q^2$.*

And

Theorem 24 (Lioville)**.** *For any $\alpha \in \overline{\mathbb{Q}} \cap \mathbb{R}$, but $\alpha \notin \mathbb{Q}$, $|\alpha - p/q| \geq C(\alpha)q^{-d}$, and the constants involved are effectively computable.*

Baker's theory gives us an improvement over Lioville, giving $q^{-d-d(\alpha)}$ for some $d(\alpha) > 0$ effectively. This is a significant advance for effectively solving Thue equations, etc. However, the main theorem of the entire subject is

Theorem 25 (Roth, 1950s)**.** *For any $\alpha \in \overline{\mathbb{Q}} \cap \mathbb{R}$, but $\alpha \notin \mathbb{Q}$,*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{C(\alpha, \delta)}{q^{2+\delta}}$$

for any $\delta > 0$.

Roth's theorem is great, but completely ineffective. From next week onwards, we'll work on the proof of Roth's theorem. But today, we'll do Thue's result from around 1910 which got the entire subject started, and achieves a bound of $q^{-(n/2+\delta)}$, ineffectively. We also have

Conjecture 7 (Lang)**.** *For any $\alpha \in \overline{\mathbb{Q}} \cap \mathbb{R}$, but $\alpha \notin \mathbb{Q}$,*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{C(\alpha, \kappa)}{q^2 (\log q)^\kappa}$$

if κ is sufficiently large.

Braver still is that we can take $\kappa > 1$. $\kappa = 1$ would of course not work, recall a popular question on the qualifying exam in measure theory.

Anyway, as we were saying, we have a result of Thue from 1909, which says that $|\alpha - p/q| \geq \frac{C(\alpha,\eta)}{q^{n/2+1+\eta}}$ for any $\eta > 0$, ineffectively. This result was later improved by Siegel to get $q^{2\sqrt{n}+\epsilon}$, and then further refined by Gelfond and Dyson to get $q^{\sqrt{2n}+\epsilon}$, and finally finished by Roth. Recall how the Lioville bound was proven: We found a polynomial f over $\mathbb{Z}$ of which α is a root. It might seem natural to just pick the minimal polynomial for α, but to illustrate a point, let's think about f possibly being larger than just the minimal polynomial. If p/q is an approximation to α, we showed both that

- $f(p/q)$ is small, by mean value theorem.
- $f(p/q)$ is big. It is rational, so $|f(p/q)| \geq 1/q^n$.

Now, if α vanishes to order h in the minimal polynomial, $|f(p/q)| \ll |\alpha - p/q|^h$, so that $|\alpha - p/q| \gg \frac{1}{q^{n/h}}$. But $\deg f \geq hd$, so you don't gain anything by picking a polynomial larger than the minimal polynomial. Thue's idea is that although we cannot exploit going to a higher degree polynomial directly, if we go to a polynomial in two variables, such a change becomes significant.

So, Thue's idea: Let $F(x, y) = P(x) - yQ(x)$, where P, Q are polynomials of degree $\leq k$ with integer coefficients. We want to construct $F(x, \alpha)$ to vanish at $x = \alpha$ to order h. We pick p_1/q_1, and p_2/q_2 to be two good rational approximations to α. We will be able to rig things so that $F(\frac{p_1}{q_1}, \frac{p_2}{q_2})$ is very small. Also, we'll use a trivial lower bound for $F(\frac{p_1}{q_1}, \frac{p_2}{q_2})$. If it's not zero, then we'll be

ok. We'll get around this nonzero problem by taking $F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2})$ small, where t is some small derivative in x. That is, let

$$F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2}) := \frac{1}{t!}\frac{d^t}{dx^t}F(x,y) \in \mathbb{Z}[x].$$

Assume α is an algebraic integer.

The first step is to construct P, Q with coefficients of small size. We have $2k$ free variables, and want $F_j(\alpha, \alpha) = 0$ for all $0 \leq j \leq h-1$. Let the height of α be B. now, α^j can be written as a linear combination of $1, \alpha, \alpha^2, \ldots, \alpha^{d-1}$ by repeatedly reducing. The coefficients in this linear combination will be of size B^j. So we have hd equations like $F_j(\alpha, \alpha) = 0$ in $2k$ variables. The size of the coefficients is $\leq B^k \frac{k^j}{j!}$, where the first factor is from the α's and the second is from differentiating. So this is just $\leq B^k e^k$. So pick $k = \frac{hd}{2}(1+\delta)$ for some small $\delta > 0$. By Thue-Siegel, there exists P, Q, polynomials with these properties and the coefficients of P and Q are $\leq (Ck(Be)^k)^{2k} 2k - hd \leq c^k$ for some $c = c(\alpha, \delta)$.

The second step is to obtain an upper bound for $|F(\frac{p_1}{q_1}, \frac{p_2}{q_2})|$. We have

$$F(\frac{p_1}{q_1}, \frac{p_2}{q_2}) = F(\frac{p_1}{q_1}, \alpha) + (\alpha - \frac{p_2}{q_2})Q(\frac{p_1}{q_1}) \leq |F(\frac{p_1}{q_1}, \alpha)| + C^k|\alpha - \frac{p_2}{q_2}|.$$

Now, for the t-th derivative, we have the same thing:

$$\begin{aligned}
F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2}) &= F_t(\frac{p_1}{q_1}, \alpha) + (\alpha - \frac{p_2}{q_2})Q_t(\frac{p_1}{q_1}) \\
&\leq |F_t(\frac{p_1}{q_1}, \alpha)| + C^k|\alpha - \frac{p_2}{q_2}|; \\
F_t(\frac{p_1}{q_1}, \alpha) &= \sum_j \binom{t+j}{j} F_{t+j}(\alpha, \alpha)(\frac{p_1}{p_2} - \alpha)^j; \\
|F_t(\frac{p_1}{q_1}, \alpha)| &\leq \sum_{k-t \geq j \geq h-t} \binom{t+j}{j} C^k |\alpha - \frac{p_1}{q_1}|^j \\
&\leq C^k |\alpha - \frac{p_1}{q_1}|^{h-t}; \\
|F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2})| &\leq C^k(|\alpha - \frac{p_1}{q_1}|^{h-t} + |\alpha - \frac{p_2}{q_2}|).
\end{aligned}$$

Third step: We want for some smallish t, a lower bound for $|F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2})|$. We'll go ahead and show how to finish the proof under the (possibly false!) assumption that $F(\frac{p_1}{q_1}, \frac{p_2}{q_2}) \neq 0$, and then later reduce the general argument to this case. If we knew that $F(\frac{p_1}{q_1}, \frac{p_2}{q_2}) \neq 0$, then it would be a rational number with denominator $q_1^k q_2$, i.e. $\geq \frac{1}{q_1^k q_2}$. Assume that

$$|\alpha - \frac{p_1}{q^1}| \leq \frac{1}{q_1^{n/2+1+\eta}}$$

where $\eta \to 0$ but is large compared to δ. Assume the same for p_2/q_2 satisfies the same inequality. If we could say that q_2 is bounded in terms of q_1, then we'd be done. So using the bounds established above,

$$\frac{1}{q_1^k q_2} \leq C^k \left(\frac{1}{q_1^{(n/2+1+\eta)h}} + \frac{1}{q_2^{n/2+1+\eta}} \right)$$

or say

$$\frac{1}{q_1^k q_2} \leq \frac{2C^k}{q_1^{(n/2+1+\eta)h}}$$

which gives

$$q_2 \geq 2^{-1} C^{-k} q_1^{h(1+\eta-\delta n/2)}.$$

So either this or the other inequality

$$\frac{1}{q_1^k q_2} \leq \frac{2C^k}{q_2^{n/2+1+\eta}}$$

gives

$$q_2^{n/2+\eta} \leq 2C^k q_1^k \Longrightarrow q_2 \leq C^{2k/n} q_1^{\frac{h(1+\delta)}{1+2\eta/n}}.$$

Now, if $h = \frac{\log q_2}{\log q_1}$ is sufficiently large both of our bounds on q_2 are contradicted! If $\frac{p_1}{q_1}$ exists with q_1 large enough, then there are only finitely many choices for $\frac{p_2}{q_2}$. This is where things become ineffective. We can't compute the $\frac{p_2}{q_2}$ because of course such $\frac{p_1}{q_1}$ doesn't exist! There is a paper of Bombieri where he gives classes of examples where one can make things effective (see Acta Mathematica 1981).

Now, we have to back track to showing a lower bound for $|F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2})|$ to finish the proof in general. So we want to find some small t so that $F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2}) \neq 0$. Suppose not. Then for all $t \leq T$ we have equations like

$$\begin{cases} P(\frac{p_1}{q_1}) - \frac{p_2}{q_2} Q(\frac{p_1}{q_1}) = 0 \\ P'(\frac{p_1}{q_1}) - \frac{p_2}{q_2} Q'(\frac{p_1}{q_1}) = 0, \end{cases}$$

etc. for higher derivatives. So given these first two equations, we can eliminate $\frac{p_2}{q_2}$ and get

$$P(\frac{p_1}{q_1})Q'(\frac{p_1}{q_1}) - P'(\frac{p_1}{q_1})Q(\frac{p_1}{q_1}) = 0.$$

Similarly, by picking any pair of equations corresponding to higher derivatives, we obtain

$$P^{(j)}(\frac{p_1}{q_1})Q^{(\ell)}(\frac{p_1}{q_1}) - P^{(\ell)}(\frac{p_1}{q_1})Q^{(j)}(\frac{p_1}{q_1}) = 0$$

for all $0 \leq j, \ell \leq T$. Consider the Wronskian of P, Q:

$$W(x) := \det \begin{pmatrix} P(x) & Q(x) \\ P'(x) & Q'(x) \end{pmatrix} = P(x)Q'(x) - P'(x)Q(x).$$

So this vanishes at $\frac{p_1}{q_1}$ but also to high order, i.e. many of it's derivatives vanish as well. $W \in \mathbb{Z}[x]$ of degree $\leq 2k-1$, and all of its derivatives up to $T-1$ vanish at $\frac{p_1}{q_1}$. So $W(x)$ is divisible by $(x-\frac{p_1}{q_1})^T$.

$W(x)$ is <u>not</u> identically zero:

$$\left(\frac{P(x)}{Q(x)}\right)' = 0 \Longrightarrow Q(x) = cP(x), c \in \mathbb{Q}.$$

So then $F(x,y) = P(x)(1-cy)$. And $P(x)$ vanishes to order h at $x=\alpha$. So by Gauss' lemma, $(q_1x-p_1)^T$ divides $W(x)$ as polynomials with integer coefficients. But the coefficients of $W(x) \leq C^k$, so $q_1^T \leq C^k \Longrightarrow T \leq \frac{k \log C}{\log q_1}$, so some small $t \leq T$ satisfies $F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2}) \neq 0$.

15 Roth's Theorem I

Last time we talked about Thue's theorem. It says that there are only finitely many solutions to

$$|\alpha - \frac{p}{q}| \leq \frac{1}{q^{d/2+1+\delta}},$$

where α is an algebraic number of degree d. There were three broad steps to the proof

1. Find $F(x,y) = P(x) - yQ(x)$ with small coefficients and vanishing to high order at (α, α). (Say, h is huge). We did this using Thue-Siegel.
2. $F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2})$ is small for good rational approximations $\frac{p_1}{q_1}$, $\frac{p_2}{q_2}$ to α. (Proof using Taylor). Think of q_2 much larger than q_1, and q_1 already quite large.
3. $F_t(\frac{p_1}{q_1}, \frac{p_2}{q_2}) \neq 0$ for small choice of t. This gives a lower bound for $F(\frac{p_1}{q_1}, \frac{p_2}{q_2})$

These three things contradict each other, and prove the theorem, provided choices of parameters are made appropriately. The ineffectively of Thue's result is because we assumed that we didn't have a first good approximation q_1.

Now we proceed to Roth's theorem: There are only finitely many solutions to

$$|\alpha - \frac{p}{q}| \leq \frac{1}{q^{2+\delta}}.$$

There are roughly the same three steps to the proof.

- Find $p(x_1, x_2, \ldots, x_m)$ of small degree and small coefficients vanishing to high order at $(\alpha, \alpha, \ldots, \alpha)$. Here m will depend on δ and d. The proof is again by Thue-Siegel.
- For nearby rational points $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m})$, $P(\cdots)$ is very small.
- In fact there is a lower bound for $P(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m})$.

So it's pretty complicated. For the moment, let's state some general propositions which will help us later. Without loss of generality, assume that α is an algebraic integer. $P(x_1, \ldots, x_m)$ is a polynomial with integer coefficients, and the degree in $x_j \leq r_j$.

Definition 1. *The index of P at $(\alpha_1, \ldots, \alpha_m; r_1, \ldots, r_m)$ is the smallest value of*

$$\sum_{j=1}^{m} \frac{i_j}{r_j}$$

with $P_{i_1,\ldots,i_m}(\alpha_1, \ldots, \alpha_m) \neq 0$.

Recall, we've defined

$$P_{i_1,\ldots,i_m}(x_1, \ldots, x_m) = \frac{1}{i_1! i_2! \cdots i_m!} \frac{d^{i_1+\cdots+i_m}}{dx_1^{i_1} \cdots dx_m^{i_m}} P(x_1, \ldots, x_m).$$

Proposition 2 (Construction of an auxiliary polynomial)**.** *Assume $\epsilon > 0$ is small, and $m \geq \frac{16}{\epsilon^2} \log d$. There is a $P(x_1, \ldots, x_m)$ with $\deg x_j \leq r_j$ and integer coefficients which are bounded by $B^{r_1+\cdots+r_m}$, $B = B(\alpha)$, and index of P with respect to $(\alpha, \alpha, \cdots, \alpha, r_1, \ldots, r_m)$ is at least $\frac{m}{2}(1-\epsilon)$.*

Proposition 3 (Index of P at nearby points)**.** *Take P as in the previous proposition. Let $\delta < 1$, $\delta > 36\epsilon$. Assume we have good rational approximations, with*

$$|\alpha - \frac{p_j}{q_j}| \leq \frac{1}{q_j^{2+\delta}}$$

for all j, $1 \leq j \leq m$. (In the back of our heads, we're thinking that $q_j \to \infty$ fast.) Assume that $q_j^{\delta} \geq D = D(\alpha)$ is sufficiently large, and $r_1 \log q_1 \leq r_j \log q_j \leq)(1+\epsilon) r_1 \log q_1$. Then the index of P at $\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m$ is at least ϵm.

This is like the second step in Thue's theorem. The proof is basically a Taylor series argument. The next proposition is really the key step to Roth's theorem, and also the hardest of the three propositions.

Proposition 4 (Roth's Lemma)**.** *Let $\omega = \omega(m, \epsilon)$ be $\frac{24}{2^m}\left(\frac{\epsilon}{12}\right)^{2^{m-1}}$, $\omega(1, \epsilon) = \epsilon$, then decreases pretty rapidly. Assume the r_j are rapidly decreasing, i.e. $r_j \omega \geq r_{j+1}$, and $g_j^{r_j} \geq q_1^{r_1}$, for all $j = 1, \ldots, m$, and all $q_j^{\omega} \geq 2^{3m}$, (q_j large). Then if P is a polynomial in $x_1, \ldots, x_m$ with $\deg x_j \leq r_j$, and integer coefficients $\leq q_1^{\omega r_1}$, then the index of P at $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m)$ is $< \epsilon$.*

Now, we can quickly prove Roth's theorem from these three propositions.

Proof (of Roth's theorem assuming the previous three propositions): Pick approximations $\frac{p_j}{q_j}$ to α, where $q_j \to \infty$ rapidly, then we know how to pick the r_j. Pick r_j sufficiently large, $r_j = \left(\frac{r_1 \log q_1}{\log q_j}\right) + 1$, assuming we have infinitely many good approximations to choose from. Plugging this into the propositions, we get a contradiction between propositions 3 and 4 after letting ϵ go to infinity. $\square$

So we've proven Roth after proving these three propositions. Let's go ahead and get started.

Proof (construction of aux. poly.): The number of free variables (from coefficients of the polynomials) is $\prod_{j=1}^{m}(r_j+1)$. If $i_1,\dots,i_m$ is such that $\sum \frac{i_j}{r_j} \leq \frac{m}{2}(1-\epsilon)$, we want $P_{i_1,\dots,i_m}(\alpha,\alpha,\dots,\alpha)=0$. We take

$$P(x_1,\dots,x_m)=\sum_{k_j\leq r_j} p(k_1,\dots,k_m)x_1^{k_1}\cdots x_m^{k_m}$$

We re-write the powers of α which appear above in terms of their defining equations. So if $\mathrm{ht}(\alpha)=C$, the the coefficients involved in writing α^j are $\leq C^j \leq C^{r_1+\cdots r_m}$. So each condition (choice of i's) gives d linear equations with coefficients $\leq (2C)^{r_1+\cdots+r_m}$. So the number of equations is d times the number of solutions to the index bound above. So long as the number of equations is $\leq \frac{1}{2}(r_1+1)\cdots(r_m+1)$, then Thue-Siegel would imply the proposition. So the question is does

$$\#\{0\leq i_j\leq r_j : \sum \frac{i_j}{r_j}\leq \frac{m}{2}(1-\epsilon)\} \leq \frac{1}{2d}(r_1+1)\cdots(r_m+1)?$$

Here's one heuristic idea, a probabilistic interpretation. Think of $x_j=\frac{i_j}{r_j}$ as random variables uniform on $(0,1)$. Then $\mathrm{Prob}(x_1+\cdots+x_m\leq \frac{m}{2}(1-\epsilon)) \leq e^{-c_1(\epsilon\sqrt{m})^2}$. So we would expect $x_1+\cdots+x_m$ to be Gaussian with mean $\frac{m}{2}$, and variance

$$\frac{m}{12}=\mathbb{E}((x_i-\frac{1}{2})(x_j-\frac{1}{2}))=\begin{cases}0 & \text{if } i\neq j\\ \int_0^1 (x-1/2)^2\,dx & \text{if } i=j.\end{cases}$$

So to make things work, we want to be away by $\frac{m}{2}-\sqrt{\frac{m}{12}}\left(\frac{\epsilon}{2}\sqrt{12m}\right)$ standard deviations.

But here's an actual rigorous proof, using "Rankin's Trick". Let $\lambda>0$. We

have that the number of solutions is

$$\begin{aligned}
&\leq \sum_{0\leq i_j\leq r_j} e^{\lambda\sum\frac{i_j}{r_j}+\lambda(m/2)(1-\epsilon)} \\
&= e^{\frac{\lambda\epsilon m}{2}}\prod_{j=1}^{m}\left(e^{\lambda/2}\sum_{0\leq i_j\leq r_j} e^{-\frac{\lambda i_j}{r_j}}\right) \\
&= e^{\frac{\lambda\epsilon m}{2}}\prod_{j=1}^{m}\left(\frac{\sinh\left(\frac{\lambda(r_j+1)}{2r_j}\right)}{\sinh\left(\frac{\lambda}{2r_j}\right)}\right) \\
&\leq e^{\frac{\lambda\epsilon m}{2}}\prod_{j=1}^{m}(r_j+1)\exp\left(\frac{1}{6}\sum_{j=1}^{m}\frac{\lambda^2(r_j+1)^2}{1+r_j^2}\right) \\
&\leq \left(\prod_{j=1}^{m}(r_j+1)\right)\exp\left(\frac{\lambda^2 m}{6}-\frac{\lambda m\epsilon}{2}\right)
\end{aligned}$$

Now we optimize in λ, and find that we should take $\lambda = 3\epsilon/2$. So the above is

$$\prod_{j=1}^{m}(r_j+1)\exp\left(-\frac{3}{8}m\epsilon^2\right).$$

Recall that $m = \frac{16}{\epsilon^2}\log d$, so the above reduces to

$$\leq \frac{1}{2d}\prod_{j=1}^{m}(r_j+1),$$

as was to be shown. □

Similar:

$$\psi(x,y) = \#\{n\leq x : p|n\Rightarrow p\leq y\} \leq \sum_{p|n\Rightarrow p\leq y}\left(\frac{x}{n}\right)^{\lambda} = x^{\lambda}\zeta(\lambda;y) = x^{\lambda}\prod_{p\leq y}\left(1-\frac{1}{p^{\lambda}}\right)^{-1}.$$

This bound is only a log away from the right answer.

16 Roth's Theorem II

Recall last time we proved

Proposition 5 (Construction of an auxiliary polynomial)**.** *There exists a polynomial $P(x_1,\dots,x_m)$, $m\geq\frac{16}{\epsilon^2}\log d$ which has degree in $x_j\leq r_j$, integral coefficients of size $\leq B^{r_1+\cdots+r_m}$, ($B = B(\alpha)$) and index of P at $(\alpha,\alpha,\dots,\alpha;r_1,\dots,r_m)$ is $\geq\frac{m}{2}(1-\epsilon)$.*

Now we move on to

Proposition 6 (Index of P at nearby points)**.** *Let P be as in the previous proposition. Let $0 < \delta < 1$, $0 < \epsilon < \frac{\delta}{36}$, and*

$$|\alpha - \frac{p_j}{q_j}| < \frac{1}{q_j^{2+\delta}}$$

for all $j = 1, \ldots, m$. Also, $q_j^\delta \geq D = D(\alpha)$, and $r_1, \ldots, r_m$ such that $r_1 \log q \leq r_j \log q_\leq (1+\epsilon) r_1 \log q_1$. Then the index of P at $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m)$ is at least ϵm.

Proof. $0 \leq j_1, \ldots, j_m$ with $\sum \frac{j_\ell}{r_\ell} \leq \epsilon m$, $j_\ell \leq r_\ell$.

$$Q(x_1, \ldots, x_m) = P_{j_1, \ldots, j_m}(x_1, \ldots, x_m) = \frac{1}{j_1! \cdots j_m!} \frac{d^{j_1}}{dx_1^{j_1}} \cdots \frac{d^{j_m}}{dx_m^{j_m}} P(x_1, \ldots, x_m).$$

Want $Q(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}) = 0$. Index of Q at $(\alpha, \alpha, \ldots, \alpha, r_1, \ldots, r_m)$ is at least $\geq \frac{m}{2}(1-\epsilon) - \epsilon m = \frac{m}{2}(1-3\epsilon)$. Take a Taylor expansion of Q around $(\alpha, \ldots, \alpha)$:

$$Q(\frac{p_1}{q_2}, \ldots, \frac{p_m}{q_m}) = \sum_{i_1, \ldots, i_m \geq 0} Q_{i_1, \ldots, i_m}(\alpha, \alpha, \ldots, \alpha) \left(\frac{p_1}{q_1} - \alpha\right)^{i_1} \left(\frac{p_2}{q_2} - \alpha\right)^{i_2} \cdots \left(\frac{p_m}{q_m} - \alpha\right)^{i_m}.$$

This is actually a finite sum because we took the Taylor expansion of a polynomial. If we get an upper bound for this sum, we will be able to conclude that $Q(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}) = 0$, as desired. This is because $Q(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m})$ is a rational number with denominator $q_1^{r_1} \cdots q_m^{r_m}$. We use the hypothesis of the theorem to bound the factors $\left(\frac{p_j}{q_j} - \alpha\right)^{i_j}$, and so we just need an upper bound for the derivatives of Q. The coefficients of Q are $\leq 2^{r_1 + \cdots + r_m}$. The coefficients of P are $\leq (2B)^{r_1 + \cdots + r_m}$ (because of the derivatives, e.g. from $x_\ell^{k_\ell}$ we get $\binom{k_\ell}{j_\ell}$), so coefficients of $Q_{i_1, \ldots, i_m}$ are $\leq (4B)^{r_1 + \cdots + r_m}$. Thus we find

$$|Q_{i_1, \ldots, i_m}(\alpha, \ldots, \alpha)| \leq (8B)^{r_1 + \cdots + r_m} \max(1, |\alpha|)^{r_1 + \cdots + r_m} = C^{r_1 + \cdots + r_m},$$

where $C = C(\alpha)$. Thus

$$Q(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}) \leq C^{r_1 + \cdots + r_m} \sum_{i_1, \ldots, i_m} \left(\frac{1}{q_1^{i_1} q_2^{i_2} \cdots q_m^{i_m}}\right)^{2+\delta}.$$

In fact, the sum here is only over those i_ℓ such that $\sum \frac{i_\ell}{r_\ell} \geq \frac{m}{2}(1-3\epsilon)$, because many of the derivatives vanish.

We have that $q_j^{r_j} \geq q_1^{r_1}$, so

$$(q_1^{i_1} \cdots q_m^{i_m}) \geq (q_1^{r_1})^{\frac{i_1}{r_1}} (q_1^{r_1})^{\frac{i_2}{r_2}} \cdots (q_1^{r_1})^{\frac{i_m}{r_m}} \geq (q_1^{r_1})^{\frac{m}{2}(1-3\epsilon)}.$$

We also know $q_1^{r_1} \geq q_j^{r_j/(1+\epsilon)}$, so the above is

$$\geq (q_1^{r_1} \cdots q_m^{r_m})^{\frac{1}{2}\frac{(1-3\epsilon)}{1+\epsilon}} \geq (q_1^{r_1} \cdots q_m^{r_m})^{\frac{(1-5\epsilon)}{2}}.$$

So

$$Q(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}) \leq C^{r_1+\cdots+r_m} 2^{r_1+\cdots+r_m} (q_1^{r_1} \cdots q_m^{r_m})^{-\frac{(2+\delta)(1-5\epsilon)}{2}},$$

now we win if can overcome the constant. But we know $q_j^\delta \geq$ some constant, which gives us that

$$Q(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}) \leq \frac{1}{q_1^{r_1} \cdots q_m^{r_m}},$$

thus is zero. □

Now we move on to proving the final proposition, which is the heart of the proof of Roth's theorem.

Proposition 7 (Roth's Lemma)**.** *Let $\omega = \omega(m, \epsilon) = \frac{24}{2^m} \left(\frac{\epsilon}{12}\right)^{2^{m-1}}$. Let r_j be a rapidly decreasing sequence in the sense that $r_j\omega \geq r_{j+1}$. Let $q_j^{r_j} \geq q_1^{r_1}$, $q_j^\omega \geq 2^{3m}$ is large. $P = P(x_1, \ldots, x_m)$ is a polynomial for which the degree in x_j is $\leq r_j$, and all coefficients are $\leq q_1^{\omega r_1}$. Then the index of P at $(\frac{p_1}{q_1}, \cdots, \frac{p_m}{q_m}, r_1, \ldots, r_m)$ is $\leq \epsilon$.*

Proof. Induction on m.

Base case $m = 1$. $\omega(1, \epsilon) = \epsilon$. $P(x)$ has coefficients $\leq q_1^{\epsilon r_1}$. Gauss' Lemma: $(q_1 x - p_1)^t | P(x)$ over $\mathbb{Z}[x]$ implies $t \leq \epsilon r_1$.

Induction step. Assume the lemma for $m-1$, and show it for m. Consider all expressions

$$P(x_1, \ldots, x_m) = \sum_{j=1}^{m} \phi_j(x_1, \ldots, x_{m-1})\psi_j(x_m),$$

where ϕ_j and ψ_j are polynomials over $\mathbb{Q}$. Eg. $\psi_j(x_m) = x_m^{j-1}$, $k = r_m + 1$, and we have such a decomposition. Pick such a decomposition with k minimal. We know at least that $k \leq r_m + 1$. $(\phi_1, \ldots, \phi_k)$ and $(\psi_1, \ldots, \psi_k)$ are linearly independent over $\mathbb{Q}$ or $\mathbb{R}$. E.g. $\sum c_j\phi_j = 0$; $c_k \neq 0$, $\phi_k = -\frac{1}{c_k}(c_1\phi_1 + \cdots + c_{k_1}\phi_{k-1})$.

$$P = \sum_{j=1}^{k-1} \phi_j\psi_j - \frac{1}{c_k}\sum_{j=1}^{k-1}(c_j\phi_j)\psi_k = \sum_{j=1}^{k-1} \phi_j(\psi_j - \frac{c_j}{c_k}\psi_k).$$

We now need to introduce the Generalized Wronskian. Let $f_1, \ldots, f_n$ be functions of one variable, t. The the Wronskian is

$$W(t) := \det \begin{vmatrix} f_1 & \cdots & \cdots & f_n \\ f_1' & \cdots & \cdots & f_n' \\ \vdots & & & \vdots \\ f_1^{(n-1)} & \cdots & \cdots & f_n^{(n-1)} \end{vmatrix}$$

If $f_1, \ldots, f_n$ are dependent, then the Wronskian is identically zero, but the converse is not true. But the Wronskian vanishes identically iff $f_1, \ldots, f_n$ are linearly dependent on some subinterval. Now we define the generalized Wronskian in m variables: $f_1, \ldots, f_n$ in $(x_1, \ldots, x_m)$. Consider

$$\Delta_{i_1,\ldots,i_m} := \frac{d^{i_1+\cdots+i_m}}{dx_1^{i_1}\cdots dx_m^{i_m}} \frac{1}{i_1!\cdots i_m!},$$

which is an operator of order $i_1 + \cdots + i_m$. If $f_1, \ldots, f_n$ are nice, and linearly independent, then there exists differential operators Δ_i of order at most $i-1$ such that

$$\det \begin{vmatrix} \Delta_1 f_1 & \cdots & \cdots & \Delta_1 f_n \\ \Delta_2 f_1 & \cdots & \cdots & \Delta_2 f_n \\ \vdots & & & \vdots \\ \Delta_n f_1 & \cdots & \cdots & \Delta_n f_n \end{vmatrix} \not\equiv 0,$$

i.e. the Δ_i is one of the $\Delta_{i_1,\ldots,i_m}$ defined above with $i_1 + \cdots + i_m \leq i-1$.

Now, $P(x_1, \ldots, x_m) = \sum_{j=1}^{k} \phi_j(x_1, \ldots, x_{m-1})\psi_j(x_m)$. Let

$$U(x_m) = \det\left(\frac{1}{(i-1)!}\frac{d^{i-1}}{dx_m^{i-1}}\psi_j(x_m)\right)_{1\leq i,j\leq k}$$

be a polynomial which is not identically zero. Also let

$$V(x_1, \ldots, x_{m-1}) = \det\left(\Delta_i \phi_j(x_1, \ldots, x_{m-1})\right)_{1\leq i,j\leq k},$$

and $W(x_1, \ldots, x_m)$ be the determinant of the product of these two, i.e.

$$\begin{aligned} W(x_1, \ldots, x_m) &= \det\left(\sum_{r=1}^{k} (\Delta_i \phi_j(x_1, \ldots, x_{m-1}))\frac{1}{(j-1)!}\frac{d^{j-1}}{dx_m^{j-1}}\psi_r(x_m)\right)_{1\leq i,j\leq k} \\ &= V(x_1, \ldots, x_{m-1})U(x_m) \\ &= \det\left(\Delta_i \frac{1}{(j-1)!}\frac{d^{j-1}}{dx_m^{j-1}} P(x_1, \ldots, x_m)\right)_{1\leq i,j\leq k}. \end{aligned}$$

This is a polynomial in $x_1, \ldots, x_m$ over $\mathbb{Z}$. It is not identically zero because U, V weren't. Let θ be the index of P at $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m)$, and λ be the index of W at $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m)$. We will use the inductive hypothesis of Roth's theorem to prove

Lemma 5.

$$\lambda \leq \frac{k\epsilon^2}{6}.$$

Then one can get a lower bound for λ in terms of θ and this will complete the proof.

Remark: We have relations with the index function, $\mathrm{Ind}(P_1P_2) = \mathrm{Ind}(P_1) + \mathrm{Ind}(P_2)$, and $\mathrm{Ind}(P_1 + P_2) \geq \min(\mathrm{Ind}(P_1), \mathrm{Ind}(P_2))$. □

17 Roth's Theorem III

We need to prove Roth's lemma. We have the following relations among our parameters:

- $\omega = \omega(m, \epsilon) = \frac{24}{2^m} \left(\frac{\epsilon}{12}\right)^{2^{m-1}}$.
- The r_i are rapidly decreasing: $\omega r_j \geq r_{j+1}$.
- $q_j^{r_j} \geq q_1^{r_1}$.
- $q_j^{\omega} \geq 2^{3m}$.

Let P be a polynomial with integral coefficients in the variables $(x_1, \ldots, x_m)$, where the degree in x_j is $\leq r_j$, and the coefficients are $\leq q_1^{\omega r_1}$. Then the index of P at $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}, r_1, \ldots, r_m)$ is at most ϵ.

Morally speaking, this lemma says that a polynomial with small integer coefficients cannot vanish to high order.

Proof. By induction. The case $m = 1$ followed by Gauss' Lemma.

We were working on the induction step $m - 1 \to m$, with $m \geq 2$. Idea: we peel off the last coefficient:

$$P(x_1, \ldots, x_m) = \sum_{j=1}^{k} \phi_j(x_1, \ldots, x_{m-1}) \psi_j(x_m)$$

which is a factorization over $\mathbb{Q}$. E.g. $\psi_j(x_m) = x_m^{j-1}$, writing it this way we have $k \leq r_m + 1$. Of all possible decompositions of this type, we choose one with k minimal. Then $\phi_1, \ldots, \phi_k$ are linearly independent and $\psi_1, \ldots, \psi_k$ are linearly independent over $\mathbb{R}$ (using minimality). Then we let

$$U(x_m) = \det\left(\frac{1}{(i-1)!} \frac{d^{i-1}}{dx_m^{i-1}} \psi_j\right)_{1 \leq i,j \leq k},$$

and it is not identically zero. ϕ_i are linearly independent over $\mathbb{R}$ implies that there is some generalized Wronskian

$$V(x_1, \ldots, x_{m-1}) = \det(\Delta_i \phi_j)_{1 \leq i,j \leq k},$$

where Δ_i are differential operators of order $\leq i - 1$. We can choose such a generalized Wronskian which is not identically zero.

Now we put these two Wronskians together and define:

$$\begin{aligned} & W(x_1, \ldots, x_m) \\ = \; & V(x_1, \ldots, x_{m-1}) U(x_m) \\ = \; & \det\left(\sum_{r=1}^{k} \Delta_i \phi_r \frac{1}{(j-1)!} \frac{d^{j-1}}{dx_m^{j-1}} \psi_r\right) \\ = \; & \det\left(\Delta_i \frac{1}{(j-1)!} \frac{d^{j-1}}{dx_m^{j-1}} P\right) \end{aligned}$$

Let θ be the index of P. Our goal is to prove $\theta \leq \epsilon$. Let λ be the index of W. If θ is large then λ is large. In other words, we'll get a bound for λ in terms of θ. We'll show the lemma that $\lambda \leq \frac{k\epsilon^2}{6}$. This is where the induction hypothesis will be used. It is easy to see that

$$\mathrm{Ind}(P_1P_2) = \mathrm{Ind}(P_1) + \mathrm{Ind}(P_2)$$

and

$$\mathrm{Ind}(P_1 + P_2) \geq \min(\mathrm{Ind}(P_1), \mathrm{Ind}(P_2)).$$

Now, let's deduce Roth's lemma from this sublemma. So, we derive the bound on θ from the bound on λ. The defining determinant for W is a sum of $k!$ terms. The index of $\Delta_i \frac{d^{j-1}}{dx_m^{j-1}} P$ is $\geq \theta - \frac{i_1}{r_1} - \cdots - \frac{i_{m-1}}{r_{m-1}} - \frac{(j-1)}{r_m}$. Recall definition of Δ_i, derivatives in each variable to orders $i_1, \ldots, i_{m-1}$, with $\sum i_\ell \leq i-1$. In fact,

$$\begin{aligned}
\mathrm{Ind}(\Delta_i \frac{d^{j-1}}{dx_m^{j-1}} P) &\geq \theta - \frac{i_1}{r_1} - \cdots - \frac{i_{m-1}}{r_{m-1}} - \frac{(j-1)}{r_m} \\
&\geq \theta - \frac{(i_1 + \cdots + i_{m-1})}{r_m - 1} - \frac{(j-1)}{r_m} \\
&\geq \theta - \frac{k-1}{r_{m-1}} - \frac{(j-1)}{r_m} \\
&\geq \theta - \frac{r_m}{r_{m-1}} - \frac{(j-1)}{r_m} \\
&\geq \theta - \omega - \frac{(j-1)}{r_m}.
\end{aligned}$$

And ω is at most ϵ^2. The index is always ≥ 0, hence

$$\lambda = \mathrm{Ind}(w) \geq \sum_{j=1}^{k} \max(0, \theta - \omega - \frac{(j-1)}{r_m}),$$

which implies that

$$\lambda + \omega k \geq \sum_{j=1}^{k} \max(0, \theta - \frac{j-1}{r_m}).$$

So at this point, we've justified (with some computations) our claim that there is an upper bound for θ in terms of λ. So, using the lemma, $\frac{k\epsilon^2}{4} \geq \lambda + k\omega$. When is this better than just using 0?

Case 1 $\theta \geq \frac{k-1}{r_m}$. Then $\frac{\theta k}{2} \leq \theta k - \frac{k(k-1)}{2r_m} \leq \frac{k\epsilon^2}{4}$, which implies that $\theta \leq \frac{\epsilon^2}{2} < \epsilon$.

Case 2 $\theta \leq \frac{k-1}{r_m}$. Then $\frac{k\epsilon^2}{4} \geq \sum_{j \leq \theta r_m + 1} (\theta - \frac{j-1}{r_m}) = \theta(\lfloor \theta r_m \rfloor + 1) - \frac{(\lfloor \theta r_m + 1 \rfloor)\lfloor \theta r_m \rfloor}{2r_m} \geq \frac{\theta}{2}(\lfloor \theta r_m \rfloor + 1) \geq \frac{\theta^2 r_m}{2}$. This implies $\theta \leq \epsilon\sqrt{\frac{k}{2r_m}} \leq \epsilon$. Recall that $k \leq r_m + 1$.

There are two more things we need to do to complete the proof: prove the lemma and prove some results we are using about Wronskians. Let's prove the lemma.

Proof (of sublemma): We can assume that we have a factorization of the form $W = U^*V^*$ with U^* and V^* having integer coefficients. Now we bound the coefficients of W. The coefficients of

$$\Delta_i \frac{1}{(j-1)!} \frac{d^{j-1}}{dx_m^{j-1}} P(x_1, \ldots, x_m)$$

are

$$\leq q_1^{\omega r_1} 2^{r_1 + \cdots + r_m}$$

and the number of monomials in this is

$$\leq (r_1 + 1) \cdots (r_m + 1) \leq 2^{r_1 + \cdots r_m}$$

so the coefficients of W are

$$\leq k!(\text{terms in product})(4^{r_1+\cdots+r_m} q_1^{\omega r_1})^k \leq (2^{3mr_1} q_1^{\omega r_1})^k \leq q_1^{2\omega r_1 k}.$$

The the coefficients of U^* and V^* are $\leq$ the coefficients of W, which are $\leq q_1^{2\omega r_1 k}$. So

$$\text{Ind}(U^*(x_m)) \leq \frac{k\epsilon^2}{12},$$

so

$$(q_m x_m - p_m)^t | U^*(x_m)$$

in $\mathbb{Z}[x_m]$. This in turn implies that

$$q_m^t \leq q_1^{2\omega r_1 k} \Rightarrow t = \frac{2\omega r_1 k \log q_1}{\log q_m} \leq 2\omega k r_m \Rightarrow \text{Index} = \frac{t}{r_m} \leq 2\omega k \leq \frac{k\epsilon^2}{12}.$$

□

But we can bound the index of V^* in the same way, using the induction hypothesis for $m-1$ variables. Take $\omega(m-1, \frac{\epsilon^2}{12}) = 2\omega(m, \epsilon)$, so our bound for coefficients of W is $\leq q_1^{2\omega r_1 k} = q^{\omega(m-1, \frac{\epsilon^2}{12})}$. Take $\epsilon \to \frac{\epsilon^2}{12}$, $r_1 \to kr_1, \ldots, r_{m-1} \to kr_{m-1}$. So the hypotheses of Roth's Lemma are met with these replacements. So Roth's Lemma gives that

$$\text{Ind}(V^* \text{ at } (\frac{p_1}{q_1}, \ldots, \frac{p_{m-1}}{q_{m-1}}, kr_1, \ldots, kr_{m-1})) \leq \frac{\epsilon^2}{12}$$

hence

$$\text{Ind}(V^* \text{ at } (\frac{p_1}{q_1}, \ldots, \frac{p_{m-1}}{q_{m-1}}, r_1, \ldots, r_{m-1})) \leq \frac{k\epsilon^2}{12}.$$

(We could have done the same for U^*, but we worked it out explicitly instead). So we've proven Roth's Lemma. □

Also, there's Falting's product theorem, which is an improvement of Roth's lemma. See also the recent article by Bostan and Dumes in AMM Oct 2010. We should still say something about Wronskains. We want to say something like "If $f_1, \ldots, f_n$ are in one variable, t, and $\det\left(\frac{d^{i-1}}{dx^{i-1}} f_j\right)_{1\le i,j\le n} \equiv 0$ then the $f_1, \ldots, f_n$ are linearly dependent". But this isn't quite possible, as the following example illustrates:

Example 2 (Peano)**.** *Let $f_1(t) = t^2$, and $f_2(t) = t|t|$. Then the Wronskian of these two functions is $\equiv 0$, but f_1 and f_2 are linearly independent over $\mathbb{R}$.*

But, of course, they are dependent functions on $(0, \infty)$ or $(-\infty, 0)$. It turns out that this is as bad as things can ever get. We'll show that if the Wronskian is $\equiv 0$ on an interval, then the functions are linearly dependent on a subinterval thereof.

Let $f_1, \ldots, f_n \in K[[t]]$.

1. $f_i = a_i t^{d_i}$, a_i not zero, $d_1, \ldots, d_n$ distinct, Wronskian $\not\equiv 0$.

$$\det \begin{pmatrix} a_1 t^{d_1} & \cdots & a_n t^{d_n} \\ a_1 d_1 t^{d_1 - 1} & \cdots & a_n d_n t^{d_n - 1} \\ & \vdots & \end{pmatrix} = (a_1 \cdots a_n) t^{d_1 + \cdots + d_n - \frac{n(n-1)}{2}}.$$

   This is basically just a Vandermonde determinant.

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ d_1 & d_2 & \cdots & d_n \\ d_1(d_1-1) & d_2(d_2-1) & \cdots & d_n(d_n-1) \\ & & \vdots & \end{pmatrix}$$

   So use the result on Vandermondes.

2. f_i find binomial of least degree. Assume $f_1 = a_1 t^{d_1} + \cdots$, $f_2 = a_2 t^{d_2} + \cdots, \ldots$, with $d_1 < d_2 < \cdots < d_n$. Strict inequality by row operations.

Next class: finish the m variable case: $f_1(x_1, \ldots, x_m), \ldots, f_n(x_1, \ldots, x_m)$. $x_1 = t$, $x_2 = t^d$, $x_3 = t^{d^2}, \ldots, x_m = t^{d^{m-1}}$. d very large. Apply the result for the one variable case.

18 Wronskians, p-adic Roth, Applications

Wronskians. Let $f_1, \ldots, f_n \in K[[t]]$. If $f_1, \ldots, f_n$ are linearly independent, then $\det\left(\frac{d^{i-1}}{dt^{i-1}} f_j\right)_{1\le i,j\le n} \not\equiv 0$. Now, let $f_1, \ldots, f_n$ in $x_1, \ldots, x_m$, f_j polynomials. If f_j are linearly independent, then some generalized Wronskian is $\not\equiv 0$.

To show this, we reduce to the one variable case. Let d be large so that $(d-1)$ exceeds the degrees of x_j for all polynomials. $x_1 = t, x_2 = t^d, x_3 =$

$t^{d^2}, \ldots, x_m = t^{d^{m-1}}$. And $x_1^{a_1} \cdots x_m^{a_m} = t^{a_1 + a_2 d + \cdots + a_m d^{m-1}}$. $a_j \leq d-1$ so that distinct monomials give distinct powers of t. Let

$$F_j(t) = f_j(t, t^d, \ldots, t^{d^{m-1}}).$$

If f_j are linearly independent, then F_j is linearly independent. So then

$$\det\left(\frac{d^{i-1}}{dt^{i-1}} F_j\right)_{1 \leq i,j \leq n} \not\equiv 0,$$

$$\frac{d}{dt} F_j = \frac{d}{dt} f_j(t, t^d, \ldots, t^{d^{m-1}}) = \frac{d}{dx_1} f_j(t, \ldots, t^{d^{m-1}}) + (dt^{d-1}) \frac{d}{dx_2}(f_j(\cdots)) + \cdots$$

so

$$\frac{d^{i-1}}{dt^{i-1}} F_j = \text{linear combination of } \Delta_{i_1, \ldots, i_m} f_j(x_1, \ldots, x_m)|_{(t, t^d, \ldots, t^{d^{m-1}})}$$

where the coefficients are fixed polynomials in t. Now that we've proven the necessary fact about Wronskians, the proof of Roth's theorem is complete.

A quick review of the proof of Roth's theorem would be:

1. There exists $P(x_1, \ldots, x_m)$, of degrees $r_1, \ldots, r_m$ with large index at $(x_1, \ldots, x_m, r_1, \ldots, r_m)$. The proof was by Thue-Siegel and is very general.

2. If we have a polynomial of this type and if $\frac{p_j}{q_j}$ are very good approximations to α, then the index at $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m})$ is $\gg \epsilon m$. (Take Taylor expansion, many of the first terms vanish, so it's a very good approximation.) $P_{i_1, \ldots, i_m}(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m})$ estimate using Taylor approximations. $\geq \frac{1}{q_1^{r_1} \cdots q_m^{r_m}}$ if not zero.

3. $P(x_1, \ldots, x_m)$; coefficients are small, $\omega r_j \geq r_{j+1}$. Index at $(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m})$ is $\leq \epsilon$.

These three things are contradictory. The proof of 3. is quite involved. Recapitulation of proof: Pick $P(x_1, \ldots, x_m) = \sum_{j=1}^{k} \phi_j(x_1, \ldots, x_m)\psi_j(x_m)$ with k minimal, $\phi_1, \ldots, \phi_k$ linearly independent, $\psi_1, \ldots, \psi_k$. Construct out of this some generalized Wronskian which is $\not\equiv 0$. Can still maintain control on the coefficients of W.

$$W(x_1, \ldots, x_m) = V(x_1, \ldots, x_{m-1})U(x_m) = \det\left(\Delta_i \frac{1}{(j-1)!} \frac{d^{j-1}}{dx_m^{j-1}} P(x_1, \ldots, x_m\right)_{i,j}.$$

So control of the coefficients of W requires control of the coefficients of U, V. Induction on Roth's lemma shows that the index of U, V is small. So the index of W is small, which implies that the index of P is small. This last implication involves taking like a square root, $\epsilon^2 \to \epsilon$, so this is why we needed $\omega \approx \epsilon^{2^m}$. We take the square root because when you take a derivative, you lose something on the index. Then add it all up. One way to think about this proof is

that $P(x_1, \ldots, x_m)$ has some crazy singularity at $\alpha, \alpha, \ldots, \alpha$, but going to the Wronskian resolves exactly what the singularity looks like.

Now we want a p-adic version of Roth. Facts 1. and 3. don't refer to the archemedian place, so they don't change. However, fact 2. has to change, but the revision won't take too much effort. The original proof mostly goes through, so we'll get a p-adic Roth's theorem by the same general proof. We can also handle several places at once.

- Mahler: p-adic Diophantine approximation.
- Ridout: p-adic version of Roth
- LeVeque: Roth for number fields, i.e. $|\alpha - \beta|$, where$\beta \in K$, in terms of the height of β.

Together, Ridout and LeVeque's results imply the following theorem of Lang:

Theorem 26 (Lang)**.** *Let K be a number field. Let S be a finite set of places in K. For each $v \in S$, select an algebraic number α_v. Then*

$$0 < \prod_{v \in S} \min(1, |\alpha_v - \beta|_v) \leq \frac{1}{H(\beta)^{2+\delta}}$$

has only finitely many solutions for any $\delta > 0$.

Where $|p|_p = \frac{1}{p}$, $|x|_p = \prod_{v|p} |x|_v$, and $H(\beta) := \prod_{v \in K} \max(1, |\beta|_v)$ is a new height function which we'll talk about in more detail next week. It's sometimes called the "Mahler measure for β". It has very interesting properties. For example,

$$H(\zeta) = 1 \Leftrightarrow \zeta \text{is a root of unity.}$$

Also, note that $H(\frac{p}{q}) = \max(|p|, |q|)$. Another interesting fact: this height admits dealing with $\alpha_j = \infty$. What is $\infty - \beta$? If we define $\infty - \beta := \beta^{-1}$, everything works correctly.

Corollary 9. *Let α be a real algebraic number. Take a decimal expansion: $0.x_1x_2x_3\ldots$. For every n, let $\ell(n)$ be the smallest number such that $x_{n+\ell(n)} \neq 0$. À priori, Roth gives that $|\alpha - \frac{(\cdots)}{10^n}| > \frac{1}{10^{(2+\epsilon)n}}$, so $\ell(n) \leq (1+\epsilon)n$. In fact, $\ell(n) = o(n)$.*

Proof. Let $S = \{\infty, 2, 5\}$, $\alpha_\infty = \alpha$, $\alpha_2 = \infty$, $\alpha_5 = \infty$. So

$$\left|\alpha_2 - \frac{(\cdots)}{10^n}\right|_2 = \left|\frac{10^n}{(\cdots)}\right|_2 = 2^{-n},$$

so

$$\left|\alpha - \frac{(\cdots)}{10^n}\right| \frac{1}{2^n} \frac{1}{5^n} \leq \frac{1}{10^{n(2+\delta)}}.$$

Which shows that eventually, $\ell(n) \leq \delta n$ for any δ. □

Another use of Roth's theorem: Application to S-unit equation. $u+v=1$, K a number field, S a finite set of places including all of the infinite places. Then $u+v=1$ has only finitely many solutions (effectively!) This was a corollary of Baker's theorem. Now Roth gives an ineffective way of solving the S-unit equation, but its proof is instructive, and gives a bound for the number of solutions (Evertse's theorem).

Proof (of finitely many solutions): Take m an integer large compared to $s=|S|$. If $u+v=1$, we get a finite number of equations $\alpha x^m+\beta y^m=1$, with x,y S-units. If $u+v=1$ has infinitely many solutions, then one of these also has infinitely many solutions, by pigeonhole principle. So for that α,β,

$$\left(\frac{x}{y}\right)^m+\frac{\beta}{\alpha}=\frac{1}{\alpha y^m}.$$

So want to say that we can find a solution with y large in some valuation in our set. Let $w\in S$ be such that $\max_{v\in S}|y|_v=|y|_w$. So for a fixed $w\in S$, and α,β, there are infinitely many solutions to

$$\frac{\beta}{\alpha}+\left(\frac{x}{y}\right)^m=\frac{1}{\alpha y^m}.$$

The left hand side is

$$=\prod_{\zeta^m=1}\left(\frac{x}{y}-\zeta_m\sqrt[m]{\frac{-\beta}{\alpha}}\right),$$

so there exists ζ with

$$\left|\frac{x}{y}-\zeta_m\sqrt[m]{\frac{-\beta}{\alpha}}\right|_w\leq fracC|y|_w^m$$

So for fixed ζ,ζ',

$$\left|\frac{x}{y}-\zeta_m\sqrt[m]{\frac{-\beta}{\alpha}}\right|_w+\left|\frac{x}{y}-\zeta'_m\sqrt[m]{\frac{-\beta}{\alpha}}\right|_w\geq\left|(\zeta-\zeta')\sqrt[m]{\frac{-\beta}{\alpha}}\right|_w.$$

This shows that if one factor in the product is very small, then the rest must be large, so sufficient to take the minimal one. So if $|y|_w$ is large, we get a contradiction to Roth's theorem.

$$|y|_w\geq\left(\prod_{v\in S}\max(1,|y|_w)\right)^{1/s}=H(y)^{1/s}$$

so

$$\frac{c}{|y|_w^m}\leq\frac{C}{H(y)^{m/s}}.$$

If $m=2s+1$, contradiction. □

A good place to look for a proof of the p-adic Roth's theorem: Hindry and Silverman: Diophantine Geometry, part D.

A theorem which has become quite fashionable in the past 10 years is the Schmidt subspace theorem. Here is an application due to Bugeowd, Corvaj and Zannier: $\gcd(2^n - 1, 3^n - 1) \leq \exp(\epsilon n)$ if n large, for any $\epsilon > 0$.

Here's another application due to Stewart (recently posted to the arxiv): Let $P(m)$ be the largest prime dividing m. Then

$$\frac{P(2^n - 1)}{n} \to \infty$$

as $n \to \infty$.

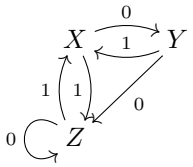
Finally, a result of Adamczewski and Bugeowd says that any irrational number which comes from a finite automaton is transcendental.

Take $x_1, x_2, \ldots$ with values in $0 \leq x_j \leq b - 1$. Look at any word length n. Is this a subword of $x_1 x_2 \cdots$? Let $\rho(n)$ be the number of length n words that appear as subwords. Then the theorem of Adamczewski and Bugeowd says that if α is a real algebraic number, and we write its base b expansion, then

$$\lim_{n \to \infty} \frac{\rho(n)}{n} = \infty.$$

This doesn't prove normalness, but says it is complemented.

A finite automaton is a sort of algorithm which a binary number as it's input, and another binary number as output. An example of a finite automation is



Here's how it works. Say we have a string of 0s and 1s which we are to read in to this machine. We assign an output value to each state, say, X outputs 1, Y outputs 1, and Z outputs 0. And say the machine starts in state X. Then if we read in the digits 01101..., say, then the machine moves to states Y, X, Z, Z, X, and hence outputs 11001...

19 The Subspace Theorem; Mahler Measure

Last time we discussed applications of the subspace theorem, but we didn't actually say what it is. Roth's theorem can be viewed as a special case of the subspace theorem. In the language of the subspace theorem, Roth's theorem would be

Theorem 27 (Roth's Theorem, subspace version)**.** *Let $L_1(x, y)$ and $L_2(x, y)$ be two linear forms in two variables, linearly independent over $\mathbb{Q}$, and with*

algebraic coefficients. Then

$$|L_1(x_1, x_2)L_2(x_1, x_2)| \leq (\max(|x_1|, |x_2|))^{-\delta}$$

for $\delta \geq 0$ has only finitely many solutions.

So, if we take $L_2(x_1, x_2) = x_2$, and $L_1(x_1, x_2) = x_1 - \alpha x_2$, and $\alpha \in \overline{\mathbb{Q}}\backslash\mathbb{Q}$, we recover Roth's theorem. If L_1 is small, then L_2 is large.

Theorem 28 (Subspace Theorem (Schmidt))**.** *Let $L_1, \ldots, L_n$ be linear forms in $x_1, \ldots, x_n$ with coefficients in $\overline{\mathbb{Q}}$, linearly independent over $\mathbb{Q}$. Then if*

$$|L_1(x_1, \ldots, x_n) \cdots L_n(x_1, \ldots, x_n)| \leq (\max |x_1|, \ldots, |x_n|)^{-\delta}$$

for any $\delta > 0$ then the solutions $(x_1, \ldots, x_n)$ are contained in finitely many proper subspaces of $\mathbb{Q}^n$.

Here is an example to show that subspaces are actually necessary. Consider the following linear forms:

$$L_1 = x_1 + \sqrt{2}x_2 + \sqrt{3}x_3$$

$$L_2 = x_1 + \sqrt{2}x_2 - \sqrt{3}x_3$$

$$L_3 = x_1 - \sqrt{2}x_2 - \sqrt{3}x_3.$$

For the purposes of this example, we take the subspace of $\mathbb{Q}^3$ defined by $x_3 = 0$. Then, consider the infinitely many solutions to Pell's equation $x_1^2 - 2x_2^2 = 1$, say with $x_1 > 0$ and $x_2 < 0$ so that $x_1 + \sqrt{2}x_2$ is small. Then for any one of these infinitely many solutions,

$$|L_1L_2L_3| = |x_1 + \sqrt{2}x_2| \leq \frac{1}{x_1 - \sqrt{2}x_2} \leq (\max |x_1|, |x_2|)^{-\delta},$$

for many choices of δ.

There is also a p-adic version of this due to Schlickewei.

Now we discuss more applications and classical extensions of Roth's theorem.

Corollary 10. *If $\alpha_1, \ldots, \alpha_k$ are algebraic with $1, \alpha_1, \ldots, \alpha_k$ linearly independent, then there are finitely many solutions to $q^{1+\delta}||q\alpha_1|| \cdots ||q\alpha_k|| \leq 1$ over $\mathbb{Q}$.*

Here $||\cdot||$ is the distance to the nearest integer function. Roth's theorem is the case $k = 1$. Here is a related famous conjecture:

Conjecture 8 (Littlewood)**.** *Given any two numbers α, β, real, prove that*

$$\liminf_{q\to\infty} q||q\alpha||||q\beta|| = 0$$

The best result we have so far is due to Lindenstrauss, who proved that the Hausdorff dimension of the set of counterexamples to this conjecture is 0.

Proof. We deduce the corollary from the subspace theorem. Let $n = k+1$, and

$$\begin{cases} L_j(x) = \alpha_j x_n - x_j & 1 \le j \le k \\ L_n(x) = x_n & j = n. \end{cases}$$

So that

$$L_1(x) \cdots L_n(x) = x_n ||x_n \alpha_1|| \cdots ||x_n \alpha_k|| \le x_n^{-\delta}$$

A solution $(x_1, \ldots, x_n)$ to the above condition lies in a proper subspace of $\mathbb{Q}^n$. Say that this subspace is defined by the equation

$$\sum_{j=1}^{n} c_j x_j = 0, \quad c_1, \ldots, c_n \in \mathbb{Q}.$$

Denote $(x_1, \ldots, x_n) = (p_1, \ldots, p_k, q)$. Then

$$c_1(\alpha_1 q - p_1) + \cdots + c_k(\alpha_k q - p_k) = (c_1\alpha_1 + \cdots + c_k\alpha_k + c_n)q \gg q.$$

Thus the q is bounded in terms of the c_j, so there are finitely many such solutions. □

Corollary 11. *Let $\alpha \in \overline{\mathbb{Q}}$. Approximate α, by all algebraic numbers β of degree $\le d$. Then there are finitely many solutions to $0 < |\alpha - \beta| \le Ht(\beta)^{-d-1-\delta}$. ($Ht(\cdot)$ denotes the naïve height.)*

Corollary 12. *If $\alpha_1, \ldots, \alpha_k$ are linearly independent and algebraic, then*

$$|x_1\alpha_1 + \cdots + x_k\alpha_k| \le (\max|x_1|, \ldots, |x_k|)^{-k+1-\delta}$$

has finitely many solutions.

The above corollary again recovers Roth's theorem. The proof uses a pigeonhole argument, assuming there are infinitely many solutions with $(\cdots)^{-k+1}$. For these facts, see the article by Bilu in the Seminaire Bourbaki.

Now we move on to discuss Mahler measure of polynomials. (N.B. The Mahler measure isn't a measure at all, but actually another height function with nice properties). Let $f(x_1, \ldots, x_n)$ be a polynomial. Then we define

$$M(f) := \exp\left(\int_0^1 \cdots \int_0^1 \log|f(e(\theta_1), \ldots, e(\theta_n))| \, d\theta_1 \cdots d\theta_n \right).$$

For example, if $f(x) = a_d x^d + \cdots + a_0 = a_d \prod_{j=1}^{d}(x - \rho_j)$, then

$$M(f) = |a_d| \prod_j \max(1, |\rho_j|).$$

For $\alpha \in \overline{\mathbb{Q}}$, we can put $M(\alpha) := M(f)$, where f is the minimal polynomial of α. We also have that

$$M(\alpha) = H(\alpha)^{\deg \alpha},$$

where $H(\cdot)$ is the absolute height defined in a previous lecture by

$$H(\alpha) = \prod_{v \in \text{places}} \max(1, |\alpha|_v),$$

where the absolute values are normalized so that $|x|_p = \prod_{v|p} |x|_v$.

Two interesting properties are: 1. If $\sigma \in Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, then $H(\sigma\alpha) = H(\alpha)$, and 2. $H(1/\alpha) = H(\alpha)$. Reference: the book of Bombieri and Gubler.

Let's think a little more about Mahler measure. How small can the Mahler measure get? It's always at least 1, which is immediate from the above. When is the Mahler measure exactly 1?

Theorem 29 (Kronecker)**.**

$$M(\alpha) = 1 \Leftrightarrow \alpha \textit{ is a root of unity}$$

Proof. We can assume without loss of generality that α is an integer and a unit. Let $\alpha = \alpha_1, \ldots, \alpha_d$ be the conjugates of α, $|\alpha_j| = 1$. For $\ell \in \mathbb{Z}$,

$$\prod_{j=1}^{d} (1 - \alpha_j^\ell) \in \mathbb{Z}$$

(it's a symmetric function). It is not zero: if it were, α would be a root of unity. So for all $\ell \neq 0$,

$$\prod_{j=1}^{d} |1 - \alpha_j^\ell| \geq 1$$

. Now, express the α_j in the form $\alpha_j = e(\theta_j)$. By Dirichlet's theorem, we find an ℓ such that $||\ell\theta_j|| \leq 1/10$ for all j, say. Contradiction. □

We can quantify the above proof. If α has degree d and $\alpha \neq$ a root of unity. Then $M(\alpha) \geq 1 + cC^{-d}$, for $c, C > 0$. Let $0 < r_j \in \mathbb{R}$, and $\alpha_j = r_j e(\theta_j)$, $\prod r_j = 1$. Then

$$M(\alpha) = \prod (1 - r_j e(\ell\theta_j)),$$

and there exists $\ell \neq 0$ with $|\ell| \leq 100^d$, and $||\ell\theta_j|| \leq 1/10$ for all j, and we can quantify exactly when Dirichlet's theorem kicks in to produce this ℓ.

There is a nicer version of this argument due to Blansky and Montgomery (1971), and they find

$$M(f) \geq 1 + \frac{1}{52d \log(6d)},$$

except if f is cyclotomic.

Here's another interesting problem concerning Mahler Measure: Consider the polynomial

$$x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1 = 0.$$

It has a unique real root α_0 with $\alpha_0 > 1$, and $1/\alpha_0 < 1$, and the remaining 8 roots are complex and on the unit circle. One computes that $M(\alpha_0) = \alpha_0 = 1.1762898$. Then there is the following

Conjecture 9 (Lehmer)**.** *α_0 is minimal with respect to Mahler measure, i.e. if $M(\alpha) < \alpha_0$, then α is a root of unity.*

Remark: Consider the unit equation $u + v = 1$, and solve it over $\mathbb{Q}(\alpha_0)$. There are finitely many solutions. In fact, there are 2532 solutions.

Smyth proved: If α is algebraic, and if $1/\alpha$ is not conjugate to α, then $M(\alpha) \geq 1.32471\ldots = \beta_0$, which is the solution to $x^3 - x - 1 = 0$, and that this is optimal. Motivated by this, we define a Pisot-Vijayaragharan number α to be a real algebraic number > 1 all of whose conjugates are < 1 in absolute value. Smyth's theorem recovers an old result of Siegel, that is, that the smallest Pisot-Vijayaragharan number is β_0.

A Salem number is defined to be a real algebraic number $\alpha > 1$ with all other conjugates ≤ 1 in size, and some conjugate on the unit circle. It is conjectured that Lehmer's example is the smallest Salem number.

Exercise 2 (Smyth)**.**

$$\log M(1 + x + y) = \frac{3\sqrt{3}}{4\pi} L(2, \chi_{-3})$$

Conjecture 10 (Deninger)**.**

$$\log M(x + \frac{1}{x} + y + \frac{1}{y} + 1) = \frac{15}{4\pi^2} L(E, 2)$$

where the L-function is normalized so that $L(E, s) = L(2 - s, E)$, and $E = x + \frac{1}{x} + y + \frac{1}{y} + 1 = 0$. Refrences: Boyd's article in Experimental Mathematics, and Rodriguez-Villegas.

20 Bilu's and Dobrowolski's Theorems

Recall last time we defined the Mahler measure of an algebraic number α. If $f(x) = a_d x^d + \cdots + a_0 \in \mathbb{Z}[x]$ is the minimal polynomial of α, then $M(\alpha) = |a_d| \prod_{j=1}^{d} \max(1, |\alpha_j|)$. There is a beautiful

Theorem 30 (Bilu)**.** *If $M(\alpha) = \exp(o(d))$, then the roots $\alpha_1, \ldots, \alpha_d$ are equidistributed around the unit circle.*

In fact, the roots actually lie in an annulus surrounding the unit circle, whose width approaches 0 as $d \to \infty$. There is also

Theorem 31 (Erdös-Turan)**.** *Let $\sum_{j=0}^{d} a_j x^j$ with small coefficients in the sense that $\sum_{j=0}^{d} |a_j| = \exp(o(d))$. Then the zeros of $f(x)$ become equidistributed as $d \to \infty$.*

Proof (sketch). In reality, all but $o(d)$ zeros satisfy $1-\epsilon \le |\alpha_j| \le 1+\epsilon$. We'll introduce a small cheat: Pretend that the zeros actually lie on the unit circle: $\alpha_j = e(\theta_j)$. This isn't actually true, but it's within epsilon of being true, so to speak, i.e. is fixable.

We have that there is an integer

$$0 \neq \operatorname{disc}(f) = a_d^{2d-2} \prod_{j \neq k} (\alpha_j - \alpha_k)$$

and

$$0 \le \log(\text{disc}) = (2d-2)\log \alpha_d + \sum_{j \neq k} \log|\alpha_j - \alpha_k| = o(d^2) + \sum_{j \neq k} \log|1 - e(\theta_j - \theta_k)|.$$

Applying the talyor expansion for log, this is

$$\le o(d^2) - \sum_{\ell \le L} \frac{1}{\ell} \sum_{j \neq k} e((\theta_j - \theta_k)\ell) + O(d^2/L)$$

$$= o(d^2) + O(d \log L) + O(d^2/L) - \sum_{\ell \le L} \frac{1}{\ell} \left| \sum_{j=1}^{d} e(\ell \theta_j) \right|^2$$

For each fixed ℓ, if

$$\sum_{j=1}^{d} e(\ell \theta_j) = o(d)$$

then the θ_j are equidistributed. Reference: Bilu: Duke Math Journal, 1997. $\square$

A particularly nice case of this circle of ideas: $x + y = 1$, x, y algebraic, and of small height has finitely many solutions. A theorem of Zagier states that apart from sixth roots of unity,

$$H(x)H(y) \ge \left(\frac{1+\sqrt{5}}{2} \right)^{1/2}.$$

There is also a theorem of Dobrowolski from 1978:

$$M(\alpha) \ge 1 + c \left(\frac{\log \log d}{\log d} \right)^3,$$

where $c = 1 - \epsilon$. Dobrowolski's theorem is a work-out of an approach first suggested by Cam Stewart using transcendence methods, so he deserves credit as well. We'll finish the course with a proof of Dobrowolski's theorem.

Proof. We can assume without loss of generality that α is not a root of unity, and that α is a unit. Let $f(x)$ be it's minimal polynomial, say of degree d. We will construct an auxiliary polynomial $F(x)$ of degree $n-1$:

$$F(x) = \sum_{j=1}^{N} a_j x^{j-1}$$

with the a_j integers. We want to keep the a_j small, and choose them so that $F(\alpha), \dots, F^{(m-1)}(\alpha)$ are all zero for some parameter M. (N, M large), i.e. $f(x)^M | F(x)$. Our plan, as usual is to use Siegel's lemma to attack this. There are Md coefficients, so .....it should be OK but there is one caveat: we must pay attention to the dependency on α. So we must make some changes to Siegel:

Lemma 6 (Siegel's Lemma Revisited)**.** *Let $b_i j \in \mathcal{O}_K$, and K be a number field of degree $d = r_1 + 2r_2$. Let*

$$\begin{cases} \sigma_1, \dots, \sigma_{r_1} & \textit{real embeddings} \\ \sigma_{r_+1}, \dots, \sigma_{r_1+r_2}, \sigma_{r_1+r_2+1}, \dots, \sigma_{r_1+2r_2} & \textit{complex embeddings} \end{cases}$$

Then, for $j = 1, \dots M$, there is a nontrivial solution to

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

in x_i with

$$|x_i| \leq Y = (2\sqrt{2}N)^{\frac{dM}{N-dM}} \left(\prod_{k,j} \max \sigma_k(b_{ij})^{\frac{1}{N-dM}} \right).$$

Proof. The proof is the same, we take the box principle and figure out what happens. Take $0 \leq y_i \leq Y$, so that the number of tuples is $= Y^N$. Now we edit a little the definition of the Galois embeddings. Let

$$\begin{cases} \tau_i = \sigma_i & \text{if } i \leq r_1 \\ \tau_{r_1+1}, \dots, \tau_{r_1+r_2} & \mathrm{Re}(\sigma_i) \\ \tau_{r_1+r_2+1}, \dots, \tau_{r_1+2r_2} & |\mathrm{Im}(\sigma_i)|. \end{cases}$$

Look at the numbers

$$\tau_k \left(\sum_{i=1}^{N} b_{ij} x_i \right) \in [-YN \max |T_k(b_{ij}), YN \max |T_k(b_{ij})],$$

and divide it into L_j boxes for each j. The number of boxes is $\prod_{j=1}^{M} L_j^d$. By box principle, we find two vectors in the same box, and after taking their difference, to find a choice of x_i for which

$$\left|\tau_k\left(\sum_{i=1}^{N} b_{ij}x_i\right)\right| \leq \frac{2YN\max_i|\tau_k(b_{ij})|}{L_j}$$

so for $k \leq r_1$,

$$\left|\sigma_k\left(\sum_{i=1}^{N} b_{ij}x_i\right)\right| \leq \frac{2YN\max_i|\sigma_k(b_{ij})|}{L_j}$$

and for $r_1 \leq k \leq r_1 + r_2$,

$$\sigma_k\left(\sum b_{ij}x_i\right)\sigma_{k+r_2}\left(\sum b_{ij}x_i\right) = \tau_k\left(\sum b_{ij}x_i\right)^2 + \tau_{k+r_2}\left(\sum b_{ij}x_i\right)^2$$

$$\leq \left(\frac{2YN}{L_j}\right)^2\left(\max_i|\tau_k(b_{ij})|^2 + \max_i|\tau_{k+r_2}(b_{ij})|^2\right)$$

$$\leq 2\left(\frac{2YN}{L_j}\right)^2 \max_i|\sigma_k\sigma_{k+r_2}(b_{ij})|$$

so that

$$N\left(\sum b_{ij}x_i\right) \leq 2^{r_2}\left(\frac{2YN}{L_j}\right)^d \prod_{k=1}^{d}\max_i|\sigma_k(b_{ij})|.$$

Now we choose

$$L_j \geq \sqrt{2}(2YN)\prod_{k=1}^{d}\max_i|\sigma_k(b_{ij})|^{1/d}$$

The constraint is then that

$$(2\sqrt{2}YN)^{dM}\prod_{k=1}^{d}\prod_{j=1}^{M}\max_i|\sigma_k(b_{ij})| \leq Y^N,$$

and we use the usual Siegel's Lemma at this stage. So then the correct choice of Y is

$$Y = (2\sqrt{2}N)^{\frac{dM}{N-dM}}\left(\prod_{k,j}\max\sigma_k(b_{ij})^{\frac{1}{N-dM}}\right)$$

as in the statement of the lemma. □

Now, back to the construction of the auxiliary polynomial:

$$F(x) = \sum_{j=1}^{N} a_j x^{j-1}$$

$$F^{(r)}(\alpha) = r!\sum a_j\binom{j-1}{r}\alpha^{j-1-r} = 0.$$

Let $b_{ir} = r!\binom{i-1}{r}\alpha^{i-1-r}$, so that for $r \le M-1$

$$\max_i |\sigma_k b_{ir}| \le N^r \max(1, |\sigma_k(\alpha)|)^N$$

so that

$$\prod_{k,r} \max_i |\sigma_k(b_{ir})| \le N^{\frac{M(M-1)d}{2}} M(\alpha)^{NM}$$

so by the revised Siegel's lemma, F exists with $|a_{ij}| \le Y$,

$$Y = \left(2\sqrt{2}N^{1+\frac{M-1}{2}} M(\alpha)^{N/d}\right)^{\frac{dM}{N-Md}},$$

with M large, $N \ge 2dM^2$; $|a_i| \le 5N^2 M(\alpha)^M \le 10N^2$. So we're happy if $M(\alpha)^M \ge 2$.

Now comes the tricky part. Let p a prime, $p \in [P, 2P]$, say. I claim that $F(\alpha^p) = 0$ in suitable ranges.

Lemma 7. *If α is an algebraic number, and not a root of unity, with conjugates $\alpha_1, \dots, \alpha_d$, then*

1. *$\alpha_i^r \ne \alpha_j^s$ for any i, j, r, s.*
2. $$\left|\prod_{i,j}(\alpha_i^p - \alpha_j)\right| \ge p^d$$

Proof. The first statement is straightforward Galois Theory. For the second,

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

and

$$\prod_j f_p(\alpha_j) = \prod_{j=1}^d pg(\alpha_j).$$

□

Now we work towards the claim preceding the lemma. If $F(\alpha^p) \ne 0$, then $\prod_{j=1}^d F(\alpha_j^p)$ is divisible by $\prod_j f(\alpha_j^p)^M$, so that

$$\left|\prod_{j=1}^d F(\alpha_j^p)\right| \ge p^d M,$$

and also the left hand side of this is

$$\le (10N^3)^d \prod_{j=1}^d (\max(1, |\alpha_j|))^{pN} = 10N^3 M(\alpha)^{pN}$$

so either these are all zero or the Mahler measure is large. So, either

$$F(\alpha^p) = 0 \quad \text{OR} \quad M(\alpha)^{pN}(10N^3)^d \geq p^{dM}$$

so that $p^M \geq N^6$ and

$$\begin{aligned} M(\alpha) &\geq \left(\frac{p^M}{10N^3}\right)^{\frac{d}{pN}} \\ &\geq \exp\left(\frac{dM}{2N}\frac{\log p}{p}\right) \end{aligned}$$

So win if this second bound is contradicted. Else, for $p \sim P$, $F(\alpha^p) = 0$. Fact: Except for $\leq \log d / \log 2$ special primes, $\deg(\alpha^p) = \deg(\alpha) = d$. F is divisible by $f_p(x)$. (True for one root, so true for all roots). But then this F is over-divisible, i.e. $N \geq \frac{dP}{\log P}$, else we're OK. Now, just have to choose N, M, p to get the optimal result. Take $N = 2dM^2$, $p^M \geq N^6$ so that $M = \frac{(const)\log d}{\log\log d}$. We get a contradiction if $\frac{p}{\log p} \geq \frac{N}{d}$, that is $P \geq 5M^2(\log M)$, or $P \sim \frac{(\log d)^2}{\log\log d}$. Then from above, we must have

$$M(\alpha) \geq \exp\left(\frac{1}{4M}\frac{\log P}{P}\right) = \exp(c\left(\frac{\log\log d}{\log d}\right)^3).$$

□