These are brief lecture notes for my course. Please send comments to: pv270@dpmms.cam.ac.uk.

1. Introduction

The course will mainly focus on two important and powerful tools in this subject, the subspace theorem and linear forms in logarithms. In this section, we state these results and discuss some basic applications, which will also allow us to compare the methods.

1.1. Diophantine Approximation. Diophantine approximation studies the approximation of real numbers with rational numbers. Since the rationals are dense inside the reals, any real number can be approximated by a rational with arbitrary precision. However, our aim is to find good approximations with rationals of small denominator.

The starting point for this subject is the following theorem of Dirichlet, which provides an upper bound for the error of the approximation that can always be achieved.

**Theorem 1** (Dirichlet). For every irrational number $\alpha$, there are infinitely many rational numbers $p/q$ such that

\[|\alpha - p/q| \leq \frac{1}{q^2}.\]

**Proof.** We fix an integer $N > 0$. We consider the image of the numbers $0, \alpha, 2\alpha, \ldots, N\alpha$ in $\mathbb{R}/\mathbb{Z}$. We identify $\mathbb{R}/\mathbb{Z}$ with $[0, 1]$, which we subdivide to $N$ disjoint intervals of length $1/N$. By Dirichlet’s box principle, there are $0 \leq n_1 < n_2 \leq N$ such that the images of $n_1\alpha$ and $n_2\alpha$ in $\mathbb{R}/\mathbb{Z}$ fall in the same interval. This means that there is an integer $p$ such that

\[|n_2\alpha - n_1\alpha - p| \leq \frac{1}{N}.\]

Dividing both sides by $q = n_2 - n_1$ and noting $q \leq N$, we see that (1) is satisfied for at least one rational $p/q$.

Now suppose that there is a finite collection of rationals $p_1/q_1, \ldots, p_k/q_k$ that satisfy (1). Then we can find another one that is distinct from each of these if we run the above argument such that $N > |\alpha - p_j/q_j|^{-1}$ for each $j$. (Here we used that $\alpha$ is irrational.)
To what extent can the approximation provided by Dirichlet’s theorem improved? This question can be studied in a large variety of settings. In this course we are mostly interested in approximating algebraic numbers. That story begins with the following result of Liouville.

**Theorem 2** (Liouville). Let \( d \in \mathbb{Z}_{\geq 2} \). For every real algebraic number \( \alpha \) of degree \( d \) over \( \mathbb{Q} \), there is constant \( c = c(\alpha) \) such that

\[
|\alpha - p/q| \geq \frac{c}{q^d}
\]

for all rational \( p/q \in \mathbb{Q} \).

**Proof.** Let \( P(x) \in \mathbb{Z}[x] \) be the minimal polynomial of \( \alpha \). Let \( p/q \in \mathbb{Q} \). We assume, as we may that \( \gcd(p,q) = 1 \) and \( |\alpha - p/q| \leq 1 \). We observe that \( P(p/q) \) is a rational number with denominator \( q^d \). Also \( P(p/q) \neq 0 \), since \( P \) is irreducible in \( \mathbb{Z}[x] \) being the minimal polynomial of \( \alpha \).

From these, we can conclude

\[
\frac{1}{q^d} \leq |P(p/q)| \leq |\alpha - p/q| \cdot \max_{x \in [\alpha-1, \alpha+1]} |P'(x)|,
\]

which proves the theorem with

\[
c = \min(1, \max_{x \in [\alpha-1, \alpha+1]} |P'(x)|^{-1}).
\]

\( \square \)

The above two proofs already contain the two most powerful tools of the subject:

- the box principle, and
- the fact that a non-zero integer has absolute value \( \geq 1 \).

Liouville’s theorem shows that for quadratic (i.e. algebraic of degree 2) irrationals, Dirichlet’s theorem is best possible apart from the precise value of the constant.

For higher degree algebraic numbers, the gap between the bounds in the theorems of Dirichlet and Liouville is much larger. The first improvement of the exponent \( d \) in Liouville’s theorem has been obtained by Thue, who replaced it by \( d/2 + 1 + \varepsilon \) for arbitrary \( \varepsilon > 0 \). This has been subsequently improved by Siegel to \( 2\sqrt{d} + \varepsilon \) and by Gelfond and Dyson independently to \( \sqrt{2d} + \varepsilon \). Finally, Roth obtained the exponent \( 2 + \varepsilon \), which is optimal up to the \( +\varepsilon \) in light of Dirichlet’s theorem.

**Theorem 3** (Roth). Let \( \alpha \) be an irrational real algebraic number. Then for all \( \varepsilon > 0 \) there is a constant \( c = c(\varepsilon, \alpha) \) such that

\[
|\alpha - p/q| \geq \frac{c}{q^{2+\varepsilon}}
\]

for all rational numbers \( p/q \).
1.2. **Thue equations.** These results are not only interesting on their own right, but also they are very useful in applications. The original motivation of Thue for improving Liouville’s estimate was the following result about a class of Diophantine equations, which became known as Thue equations.

**Theorem 4** (Thue). Let $P(x, y) \in \mathbb{Z}[x, y]$ be an irreducible homogeneous polynomial of degree at least 3, and let $m \in \mathbb{Z}$ be an integer. Then the equation

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

has only finitely many solutions in $x, y \in \mathbb{Z}$.

The assumption that $P$ is irreducible is an overkill, but something is needed to rule out degeneracies.

The connection between these equations and improvements of Liouville’s estimate should not come as a surprise. Indeed, Liouville’s theorem is essentially based on the bound $|P(x, y)| \geq 1$ (disguised in the form $|P(x/y, 1)| \geq y^{-d}$), which holds for any integers $x, y$. Then an improvement of Liouville’s theorem should be related to an improvement over $|P(x, y)| \geq 1$. We formulate this idea as follows.

**Lemma 5.** Let $P(x, y) \in \mathbb{Z}[x, y]$ be an irreducible homogeneous polynomial, and let $K \subset \mathbb{R}$ be a compact set. Then there are constants $c = c(P, K)$ and $C = C(P)$ such that for all $p/q \in \mathbb{Q}$ there is a root $\alpha$ of the polynomial $P(x, 1)$ with

$$c|P(p, q)| \leq |\alpha - p/q| \cdot q^d \leq C|P(p, q)|.$$

In light of this lemma, it is clear that Roth’s theorem (or any of its precursors discussed above) implies Thue’s theorem. In fact, we could even replace $m$ on the right hand side of the equation by any polynomial (in $\mathbb{Z}[x, y]$) of degree at most $d - 3$.

**Proof.** The inequality on the left can be obtained along the lines we have seen in the proof of Liouville’s theorem, so we turn to the other inequality.

Let

$$P(x, 1) = a \prod_{i=1}^{d}(x - \alpha_i).$$

We note that

$$P(x, y) = a \prod_{i=1}^{d}(x - \alpha_i y).$$

Suppose that $\alpha_1$ is the one nearest to $p/q$. Then there is a constant $c_0$ such that $|\alpha_i - p/q| \geq c_0$ for all $i = 2, \ldots, d$. Indeed, $c_0 = \min \frac{|\alpha_i - \alpha_1|}{2}$ would work.
Then

\[ |P(p, q)| \geq q^d \prod_{i=1}^{d} |p/q - \alpha_i| \geq q^d |p/q - \alpha_i| c_0^{d-1}. \]

The claim follows with \( C = c_0^{-d+1} \).

1.3. The subspace theorem. The subspace theorem is a far reaching generalization of Roth’s theorem proved by Schmidt. For an integer vector \((x_1, \ldots, x_n) \in \mathbb{Z}^n\), we define its height as

\[ H(x_1, \ldots, x_n) = \max_{1 \leq j \leq n} |x_j|. \]

Later in the course we will discuss heights in more detail.

**Theorem 6** (Subspace Theorem, Schmidt). Let \( n \in \mathbb{Z}_{\geq 2} \), and let \( L_1, \ldots, L_n \) be linearly independent linear forms in the variables \( x_1, \ldots, x_n \) with algebraic coefficients. Then for all \( \varepsilon > 0 \), the solutions of

\[ \prod_{j=1}^{n} |L_j(x_1, \ldots, x_n)| \leq H(x_1, \ldots, x_n)^{-\varepsilon}, \quad (x_1, \ldots, x_n) \in \mathbb{Z}^n \]

lie in a finite union of proper subspaces of \( \mathbb{Q}^n \).

To see that this is indeed a generalization of Roth’s theorem, consider the linear forms

\[ L_1(x_1, x_2) = x_1 - \alpha x_2, \quad L_2(x_1, x_2) = x_2 \]

for some real irrational algebraic number \( \alpha \). If \( |\alpha - p/q| < q^{-2-\varepsilon} \) for some \( \varepsilon > 0 \) and \( p, q \in \mathbb{Z} \), then

\[ |L_1(p, q)| \cdot |L_2(p, q)| \leq q^{-1-\varepsilon} \cdot q \leq q^{-\varepsilon} \leq C(\alpha) H(p, q)^{-\varepsilon}. \]

By adjusting \( \varepsilon \), we can eliminate the constant, and we conclude that \( (p, q) \) is contained in a finite collection of proper subspaces, which depend only on \( \alpha \) and \( \varepsilon \). This means that there is a finite collection \( B \) of rational numbers such that all \( p, q \) with \( |\alpha - p/q| < q^{-2-\varepsilon} \) must satisfy \( p = q\beta \) and hence \( p/q = \beta \) for some \( \beta \in B \). That is, \( |\alpha - p/q| < q^{-2-\varepsilon} \) holds with finitely many exceptions, and hence \( |\alpha - p/q| < Cq^{-2-\varepsilon} \) holds without exceptions with an appropriate choice of \( C \).

We make some further comments about the subspaces that arise in the conclusion of the subspace theorem.

- There are some obvious subspaces that may arise. If \( \text{Ker}(L_i) \) contains a rational subspace for some \( j \), then certainly all integer points on it will be solutions of (2).
- There are also some less obvious subspaces. For example, consider \( n = 3 \) and the linear forms

\[ L_1 = x_1 - 2^{1/2} x_2, \quad L_2 = x_1 - 2^{1/2} x_2 + x_3, \quad L_3 = x_2. \]
Now it is easy to see that a triple \((x_1, x_2, x_3) = (p, q, 0)\) is a solution of (2) if and only if 
\[ |2^{1/2} - p/q|^2 < q^{-3-\varepsilon}, \]
and there are infinitely many of those for all \(\varepsilon \leq 1\) by Dirichlet’s theorem.

- It is an instructive exercise to work out in the \(n = 3\) case which two dimensional subspaces may contain infinitely many solutions of (2). (Hint: They are all of the form \(\alpha L_i + L_j = 0\) for some \(\alpha \in \mathbb{Q}\) and indices \(i \neq j\) or contain the infinitely many solutions on a line.)

- A line (that is a 1 dimensional subspace of \(\mathbb{Q}^n\)) may contain at most finitely many solutions of (2) unless one of the \(L_j\) vanishes on it. Indeed, it is easy to see that the left hand side of (2) increases while the right hand side goes to 0 as we move to infinity along a line.

For an algebraic number \(\alpha\), we write \(f_{\alpha} \in \mathbb{Z}[x]\) for its minimal polynomial. For a polynomial \(f \in \mathbb{Z}[x]\), we write \(H(f)\) for the maximum of the absolute values of its coefficients. We call \(H(f)\) the height of \(f\).

The following application of the subspace theorem is due to Schmidt.

**Theorem 7** (Schmidt). Let \(\alpha\) be an algebraic number and let \(d \in \mathbb{Z}_{\geq 1}\). Then for all \(\varepsilon > 0\), there is a constant \(c = c(\varepsilon, \alpha, d)\) such that
\[
|\alpha - \beta| \geq \frac{c}{H(f_{\beta})^{d+1+\varepsilon}}
\]
for all algebraic numbers \(\beta \neq \alpha\) of degree \(d\).

For \(d = 1\), this is equivalent to Roth’s theorem. If \(\alpha\) is an algebraic number of degree at least \(d + 1\), then the exponent \(d + 1\) (putting the \(+\varepsilon\) aside) is optimal. However, it is not known whether for all \(\varepsilon > 0\) and transcendental \(\alpha\), there are infinitely many algebraic \(\beta\) of degree at most \(d\) such that \(|\alpha - \beta| \leq H(f_{\beta})^{-d-1+\varepsilon}\). This problem is known as Wirsing’s conjecture. For more on this subject we refer to [8, Section 3.4].

**Proof.** Fix \(\varepsilon > 0\), and let \(\beta\) be an algebraic number of degree \(d\) such that
\[
|\alpha - \beta| \leq \frac{1}{H(f_{\beta})^{d+1+\varepsilon}}.
\]
By the mean value theorem,
\[
|f_{\beta}(\alpha)| = |f'_{\beta}(\xi)(\alpha - \beta)|
\]
for some \(\xi\) in the interval whose endpoints are \(\alpha\) and \(\beta\). It follows that
\[
|f_{\beta}(\alpha)| \leq \frac{C}{H(f_{\beta})^{d+\varepsilon}}
\]
for some constant \(C\) that depends only on \(\alpha\). (Here we used that \(|\alpha - \beta| \leq 1\).)
Using the subspace theorem, we will show that if
\[ |f(\alpha)| \leq \frac{C}{H(f)^{d+\varepsilon}}, \]
for some \( f \in \mathbb{Z}[x] \) of degree \( d \), then either \( f(\alpha) = 0 \) or \( f \) belongs to a finite set that depends only on \( \varepsilon, \alpha, d \) and \( C \). Since \( f_\beta \) is irreducible (being the minimal polynomial of \( \beta \)), \( f_\beta(\alpha) = 0 \) is possible only if \( f_\beta \) is also the minimal polynomial of \( \alpha \). We can conclude that (3) may hold only for finitely many \( \beta \), and the claim of the theorem follows.

We turn to the proof of the claim we made about the polynomials \( f \) satisfying (4). Consider the linear forms

\[ L_0(x_0, \ldots, x_d) = x_0 + \alpha x_1 + \ldots + \alpha^d x_d \]

and

\[ L_j(x_0, \ldots, x_d) = x_j \]

for \( j = 1, \ldots, d \). The subspace theorem yields that the integer solutions of

\[ \prod_{j=0}^{d} |L_j(x_0, \ldots, x_d)| \leq CH(x_0, \ldots, x_d)^{-\varepsilon} \]

lie in a finite union of proper subspaces of \( \mathbb{Q}^n \). This means that the set of integer vectors \( x_0, \ldots, x_d \in [-H, H]^{d+1} \) that satisfy

\[ |L_0(x_0, \ldots, x_d)| \leq \frac{C}{H^{d+\varepsilon}} \]

lie in the union of those proper subspaces.

We would like to say more about the subspaces that arise in this way, so we will apply the subspace theorem again in these subspaces. Let \( V \subset \mathbb{Q}^{d+1} \) be a subspace of dimension at least 2 that is not contained in the \( \text{Ker}(L_0) \). We consider the restrictions of \( L_0, \ldots, L_d \) to \( V \). These span the dual of \( V \), so it contains a linearly independent subset of \( \tilde{d} := \dim(V) \) elements. Furthermore, by the basis exchange property, we can assume that \( L_0|_V \) is contained in this set. For simplicity, we assume that the restrictions of \( L_0, L_1, \ldots, L_{d-1} \) are linearly independent.

Let \( T : \mathbb{Q}^{\tilde{d}} \rightarrow V \) be a linear isomorphism such that \( V \cap \mathbb{Z}^{\tilde{d}} \subset T(\mathbb{Z}^{\tilde{d}}) \). We apply the subspace theorem to the linear forms \( L_j' := L_j \circ T \). It follows that the integer solutions of

\[ \prod_{j=0}^{d-1} |L_j'(y_0, \ldots, y_{d-1})| \leq CH(y_0, \ldots, y_{d-1})^{-\varepsilon} \]

are contained in a finite union of proper subspaces of \( \mathbb{Q}^{\tilde{d}} \). A simple calculation shows that for all integer vectors \( x_0, \ldots, x_d \in [-H, H]^{d+1} \cap \mathbb{V} \) that satisfy (5), \( (y_0, \ldots, y_{d-1}) = T^{-1}(x_0, \ldots, x_d) \) satisfies (6). This means that the set of integer vectors \( x_0, \ldots, x_d \in [-H, H]^{d+1} \cap \mathbb{V} \) that satisfy (5) lie in the finite union of proper subspaces of \( \mathbb{V} \) that the above application of the subspace theorem yields.
We apply the above argument repeatedly to subspaces of $\mathbb{Q}^{d+1}$ that arise in each application of the subspace theorem. We conclude that the set of integer vectors $x_0, \ldots, x_d \in [-H,H]^{d+1}$ that satisfy (5) lie in a finite union of subspaces, and each subspace in this collection is either 1 dimensional or is contained in $\operatorname{Ker}(L_0)$. For a 1 dimensional subspace $V \subset \mathbb{Q}^{d+1}$, there are at most finitely many integer vectors $(x_0, \ldots, x_d)$ such that 

$$|L_0(x_0, \ldots, x_d)| \leq C.$$ 

Therefore, the set of integer vectors $x_0, \ldots, x_d \in [-H,H]^{d+1}$ that satisfy (5) on which $L_0$ does not vanish belong to a finite set, which does not depend on $H$. The coefficients of a polynomial satisfying (4) satisfy (5), so the claim is proved. 

Schlickewei generalized the subspace theorem to allow $p$-adic absolute values. Similar generalizations of Roth’s theorem has been proved by Ridout and others previously. These results are especially useful in many Diophantine applications. We only state a special case, which is sufficient for our purposes. For a more general result we refer to the book of Bombieri and Gubler [7, Chapter 7].

We introduce some terminology. The places of $\mathbb{Q}$ is the set of prime numbers together with the symbol $\infty$. For each place $v$, we endow $\mathbb{Q}$ with an absolute value $| \cdot |_v$. If $v$ is a (finite) prime, then $| \cdot |_v$ is the $v$-adic absolute value. That is, for an integer $x$, we have $|x|_v = v^{-k}$, where $k$ is the largest integer so that $v^k | x$. For a rational number $x/y$, we have $|x/y|_v = |x|_v / |y|_v$. The absolute value $| \cdot |_\infty$ is the usual one, that is, $|p/q|_\infty = p/q$ if $p/q \geq 0$ and $|p/q|_\infty = -p/q$ if $p/q \leq 0$.

**Theorem 8** (S-adic subspace theorem, Schlickewei). Let $n \in \mathbb{Z}_{\geq 2}$. Let $S$ be a finite finite set of places of $\mathbb{Q}$ containing $\infty$. For each $v \in S$, let $L_{j}^{(v)}(x_1, x_2) \in \mathbb{Q}[x_1, x_2]$ be linearly independent linear forms. Then for all $\varepsilon > 0$, the solutions of

$$(7) \quad \prod_{v \in S} \prod_{j=1}^{n} |L^{(v)}_j(x_1, \ldots, x_n)|_v \leq H(x_1, \ldots, x_n)^{-\varepsilon}$$

lie in a finite union of proper subspaces of $\mathbb{Q}^n$.

Our remark about the finiteness of solutions contained in lines still stands. Now it is no longer true that the left hand side of (7) increases as we move along (integer points contained in) a line, but it will be still bounded away from 0, unless one of the forms vanishes on the line.

We finish this section with a simple application to demonstrate the power of this result. Consider the simple case $S = \{2, 3, \infty\}$, $L_{j}^{(v)}(x_1, x_2) = x_j$ for all $j$ and $v$. This will not lead us anywhere because the left hand side of (7) is always at least 1 provided $x_1, x_2 \neq 0$. However, the left hand side is rather small sometimes even for very large $x_1$ and $x_2$. E.g. if $x_1 = 2^k$ and $x_2 = 3^m$, then the left hand side will be 1. Now we can
exchange $L_2^{(\infty)}$ for $x_1 - x_2$ and obtain that a power of 2 cannot be too close to a power of 3.

A more precise statement is the following.

**Proposition 9.** For all $\varepsilon > 0$, there is $c = c(\varepsilon)$ such that

\[
|p^{2^k} - q^{3^m}| \geq c \frac{\max(2^{k}, 3^{m})^{1-\varepsilon}}{\max(p, q)}
\]

or $p^{2^k} = q^{3^m}$ for all $p, q, k, m \in \mathbb{Z}_{>0}$.

Here we wrote 2 and 3 for the sake of concreteness, but they could be replaced by any multiplicatively independent integers $\geq 2$. (Two integers are multiplicatively independent if you cannot raise each of them to some integer power and obtain the same number.)

Using the box principle, it is easy to see that this result is best possible apart from the $-\varepsilon$ in the exponent.

**Proof.** Take $S = \{2, 3, \infty\}$ and let $L_j^{(v)}(x_1, x_2) = x_j$ for all $j = 1, 2$ and $v \in S$ except that $L_2^{(\infty)}(x_1, x_2) = x_1 - x_2$. By the $S$-adic subsapce theorem, the solutions of

\[
\prod_{v \in S} \prod_{j=1}^n |L_j^{(v)}(x_1, x_2)|_v \leq H(x_1, x_2)^{-\varepsilon/2}
\]

lie in a finite union of lines.

Let $p, q, k, m \in \mathbb{Z}_{>0}$, and take $x_1 = p^{2^k}$, $x_2 = q^{3^m}$. For simplicity, we assume $3^m > 2^k$. We note that if $\max(p, q) > 3^m$, then the claim holds trivially with $c = 1$, so we assume that this is not the case. This means, in particular that $H(p^{2^k}, q^{3^m}) \leq 3^{2^m}$.

We have

\[
\begin{align*}
|L_1^{(\infty)}(p^{2^k}, q^{3^m})|_{\infty} &= p^{2^k}, & |L_2^{(\infty)}(p^{2^k}, q^{3^m})|_{\infty} &= |p^{2^k} - q^{3^m}| \\
|L_1^{(2)}(p^{2^k}, q^{3^m})|_2 &\leq 2^{-k}, & |L_2^{(2)}(p^{2^k}, q^{3^m})|_2 &\leq 1 \\
|L_1^{(3)}(p^{2^k}, q^{3^m})|_3 &\leq 1, & |L_2^{(3)}(p^{2^k}, q^{3^m})|_3 &\leq 3^{-m}.
\end{align*}
\]

Assuming that (8) does not hold with $c = 1$, we get

\[
\prod_{v \in S} \prod_{j=1}^n |L_j^{(v)}(p^{2^k}, q^{3^m})|_v \leq p^{-m} \frac{\max(2^k, 3^m)^{1-\varepsilon}}{\max(p, q)} \leq 3^{-m} \leq H(p^{2^k}, q^{3^m})^{-\varepsilon/2}.
\]

That is, (9) holds for $x_1 = p^{2^k}$, $x_2 = q^{3^m}$. In light of the first half of the proof, there are only finitely many quadruples $p, q, k, m$ such that (9) fails with $c = 1$. To conclude the proof, we set $c$ sufficiently small so that the claim holds also for this finitely many quadruples. ∎
1.4. Digit expansion of integers in two different bases. It is expected that an integer cannot have “simple” digit expansions in two multiplicatively independent bases. This vague statement can be made formal in many different ways. For example, one would expect the base 3 expansion of $2^n$ to behave like a random sequence of digits $0,1,2$. For a start, it is expected that each digit will occur if $n$ is sufficiently large. Moreover, they should occur with the same asymptotic frequency. These things are very much open and seem a long way beyond what we can prove at the moment.

Here we give a result of Senge and Strauss [21] as an application of Proposition 9. The proof we give is close to the original of Senge and Strauss, except that they used an extension of Roth’s theorem due to Lang instead of the subspace theorem.

For integers $a$ and $b \geq 2$, we write $N(a,b)$ for the number of non-zero digits in the base $b$ expansion of $a$.

**Theorem 10** (Senge, Strauss). We have

$$N(a,2) + N(a,3) \to \infty$$

as $a \to \infty$.

Again, here 2 and 3 may be replaced by a pair of multiplicatively independent integers $\geq 2$.

**Proof.** We show that for each $N \in \mathbb{Z}_{>0}$, there are only finitely many $a \in \mathbb{Z}_{>0}$ such that $N(a,2) + N(a,3) \leq N$. Let $a$ be one of these numbers. We write $\alpha_k \alpha_{k-1} \ldots \alpha_0$ and $\beta_m \beta_{m-1} \ldots \beta_0$ for the base 2 and base 3 expansions of $a$ respectively. For an interval $I \subset [0,1]$, we write

$$\alpha_I := \sum_{j: \log_a(2^j) \in I} \alpha_j 2^{j - \min(i: \log_a(2^i) \in I)}.$$

In other words, $\alpha_I$ is the integer whose base 2 expansion is the sequence of those $\alpha_j$ for which $2^j = a^t$ for some $t \in I$. We define $\beta_I$ using the base 3 expansion of $a$ in a similar manner.

By our assumption on $a$, there is some $j \in \{0, 1, \ldots, N\}$ such that

$$\alpha_{(1-3^{-j},1-3^{-j-1})} = \beta_{(1-3^{-j},1-3^{-j-1})} = 0.$$

This means that we have $a = p2^{k_j} + e_1$, where $p = \alpha_{(1-3^{-j-1},1)}, e_1 = \alpha_{(0,1-3^{-j})}$ and $k_j$ is the smallest integer so that $2^{k_j} \geq a^{1-3^{-j-1}}$. We note that

$$p \leq a/2^{k_j} \leq a^{3^{-j-1}}$$

$$e_1 \leq 2a^{1-3^{-j}}.$$

(The most important thing for us, as we will see below, is that $pe_1 < 2^{k_j/(1-\varepsilon)}$ with an appropriate $\varepsilon$ depending on $j$.) In a similar manner,
we can write \( a = q3^{m_j} + e_2 \) with
\[
q \leq a^{3^{-j-1}} \\
e_2 \leq 3a^{1-3^{-j}}.
\]

Therefore, we have
\[
|p2^{kj} - q3^{m_j}| = |e_1 - e_2| \leq 3a^{1-3^{-j}} \leq 3a^{-3^{-j}+3^{-j-1}+3^{-j-1}} \frac{2^{kj}}{\max(p, q)}.
\]

We apply Proposition 9 with some \( \varepsilon < 3^{-N} - 2 \cdot 3^{-N-1} \), and conclude that the above inequality cannot hold for arbitrarily large \( a \). This is precisely what we wanted to show. \( \square \)

This subject is also closely related to a family of problems in Ergodic theory formulated by Furstenberg the most famous of which is the following.

**Conjecture 11** (Furstenberg). If \( \mu \) is a probability measure on \( \mathbb{R}/\mathbb{Z} \) that is invariant under both maps \( x \mapsto 2x \) and \( x \mapsto 3x \), then \( \mu \) is a convex combination of Lebesgue measure and an atomic measure supported on the rationals.

Rudolph [20] proved this conjecture under an extra assumption about positivity of the entropy. This result has been reproved by many authors in a variety of ways, but no one was able to relax the entropy condition.

1.5. **Ineffectivity.** The results of Thue, Siegel, Dyson, Gelfond, Roth, Schmidt and Schlickewei we have discussed above have a key deficiency, namely that they are ineffective. What this means is that not specifying the values of the constants was not just laziness, but, in fact, it is not possible to specify them. The proofs yield the finiteness of the constants, but they do not give any information about how big they might be.

In particular, it is impossible to extract from the proof of Roth’s theorem any improvement over Liouville’s inequality for the approximation of \( 2^{1/3} \) by rational numbers with denominator less than \( 10^{10^{10^{10}}} \). And the point is that I could have written any number there. (Well, impossible is maybe too strong a word here, but see Masser [16, Chapter 12, pp144] for a very nice discussion about why proving e.g. \( |2^{1/3} - p/q| > cq^{-2.955} \) with an effective constant looks exceedingly hopeless with Thue’s method. I will be able to say more on this when we discuss the proof of Roth’s theorem.)

The situation is perhaps even worse when it comes to the applications to Diophantine equations. Solving an equation means listing all its solutions. Theorem 4 yields the finiteness of the number of solutions of the Thue equations. However, it does not reduce the problem of solving the equations to a finite search. It is possible to (effectively) bound the
number of solutions of such equations using Thue’s method. However, the proof yields no information about how large these solutions might be, so we do not know when we can stop looking solutions. Except if we find enough solutions to match the upper bound, but it is very unlikely that that many solutions exist.

This issue can be addressed in many situations using the theory of linear forms in logarithms, which we discuss next. In fact, that method often yields good enough constants so that the resulting finite search can be done not just in principle, but also in practice.

1.6. **Transcendence.** The theory of linear forms in logarithms originates in problems about transcendence. The existence of transcendental numbers was first demonstrated by Liouville using his lower bound on approximating algebraic numbers by rationals.

Indeed, consider the number

\[ \alpha = \sum_{n=1}^{\infty} 10^{-n!}. \]

If we truncate the sum after the first \(k\) terms we get a rational number with denominator \(q = 10^{-n!}\) and it approximates \(\alpha\) with error at most

\[ 2 \cdot 10^{-(n+1)!} < cq^d \]

for any fixed \(c\) and \(d\) if \(n\) is sufficiently large. Using Roth’s theorem, we can replace \(n!\) by any sequence growing like \((2+\varepsilon)^n\).

However, this method is severely limited in scope. It is known that almost all real number lack the rational approximations that would be required for proving transcendence using Roth’s theorem. In particular, it is not expected that the transcendence of classical constants like \(e, \pi\) or \(2\sqrt{2}\) would follow in this way. (However, to the best of my knowledge this has not been proved for \(\pi\) and \(2\sqrt{2}\).)

The first result about classical constants is due to Hermite, who proved the transcendence of \(e\). This has been extended by Lindemann who proved that \(e^\alpha\) is transcendental if \(\alpha\) is algebraic. This contains the transcendence of \(e\) (take \(\alpha = 1\)), and also that of \(\pi\), because \(e^{2\pi i} = 1\) is algebraic, so \(2\pi i\) cannot be. Lindemann also stated the following result (saying it follows along the same lines as the transcendence of \(\pi\)), whose proof was completed by Weierstrass.

**Theorem 12** (Lindemann, Weierstrass). For any distinct algebraic numbers \(\alpha_1, \ldots, \alpha_n\), the numbers \(e^{\alpha_1}, \ldots, e^{\alpha_n}\) are linearly independent over \(\mathbb{Q}\).

Here and everywhere in these notes, \(\mathbb{Q}\) denotes the field of algebraic numbers.

This was the state of the art at the turn of the 20’th century, and Hilbert chose the following as his 7’th problem.

**Problem 13** (Hilbert’s 7’th problem). Is \(\alpha^\beta\) always transcendental for algebraic \(\alpha \neq 0, 1\) and irrational algebraic \(\beta\)?
The meaning of $\alpha^\beta$ requires some explanation. It is understood to be $e^{\beta \log \alpha}$, where $\log \alpha$ can be taken to mean any complex number with $e^{\log \alpha} = \alpha$. Now this means that if $\beta$ is not an integer, then $\alpha^\beta$ is not uniquely determined, and the question is asked about any of the choices. On the other hand, in these notes, for $\alpha \in \mathbb{R}_{>0}$, we always mean by $\log \alpha$ its principal branch, that is $\log \alpha \in \mathbb{R}$ in this case.

Hilbert predicted that this problem would be more difficult to solve than the Riemann hypothesis, but the question was answered in the affirmative by Gelfond and Schneider independently in the mid 1930’s. This can be reformulated in terms of linear independence of logarithms.

**Theorem 14** (Gelfond, Schneider). Let $\alpha_1, \alpha_2$ be algebraic non-zero numbers and let $\log \alpha_1$ and $\log \alpha_2$ be fixed choices for their logarithms. Then $\log \alpha_1$ and $\log \alpha_2$ are linearly independent over $\mathbb{Q}$ if and only if they are linearly independent over $\overline{\mathbb{Q}}$.

*Proof of equivalence with Hilbert 7’th.* We first deduce the theorem from Hilbert 7’th problem. If $\log \alpha_1$ and $\log \alpha_2$ are linearly dependent over $\mathbb{Q}$, then there is a number $\beta \in \mathbb{Q}$ such that $\beta \log \alpha_1 = \log \alpha_2$. Exponentiating this, we get that $\alpha_1^\beta = \alpha_2$ is algebraic. So either $\beta \in \mathbb{Q}$ or $\alpha_1 = 1$. In both cases $\log \alpha_1$ and $\log \alpha_2$ are linearly dependent over $\mathbb{Q}$.

Now we prove the converse implication. Assume that $\alpha_1^\beta = \alpha_2$ is algebraic for some algebraic $\alpha_1$ and $\beta$. This means that there are choices of $\log \alpha_1$ and $\log \alpha_2$ such that $\beta \log \alpha_1 = \log \alpha_2$. By the theorem, this means that $\log \alpha_1$ and $\log \alpha_2$ are linearly dependent over $\mathbb{Q}$. So either $\log \alpha_1 = 0$ and $\alpha_1 = 1$ or $\beta \in \mathbb{Q}$. In either case, the proof is complete. □

It is natural to ask (and Gelfond did so) if the above theorem can be extended to more than 2 logarithms. This has been achieved by Baker, who proved the following result.

**Theorem 15** (Baker). Let $\log \alpha_1, \ldots, \log \alpha_n$ be logarithms of non-zero algebraic numbers and suppose they are linearly independent over $\mathbb{Q}$. Then $1, \log \alpha_1, \ldots, \log \alpha_n$ are linearly independent over $\overline{\mathbb{Q}}$.

We will prove a weaker version of this result later in the course.

This result of Baker contains the results of Hermite and Lindemann and the Gelfond Schneider theorem, (however, it does not contain the Lindemann Weierstrass theorem). Schanuel made the following general conjecture that is a (far reaching) common generalization of Baker’s theorem and that of Lindemann Weierstrass.

**Conjecture 16** (Schanuel). Let $n \in \mathbb{Z}_{\geq 1}$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ be linearly independent over $\mathbb{Q}$. Then the transcendence degree of the field

$$\mathbb{Q}(\alpha_1, \ldots, \alpha_n, e^{\alpha_1}, \ldots, e^{\alpha_n})$$

is at least $n$. 
This conjecture strengthens the Lindemann Weierstrass theorem in that for \( \mathbb{Q} \) linearly independent algebraic numbers \( \alpha_1, \ldots, \alpha_n \) it claims not only the \( \overline{\mathbb{Q}} \) linear independence of \( e^{\alpha_1}, \ldots, e^{\alpha_n} \), but also their algebraic independence. It turns out that this seemingly stronger version is equivalent to the original statement.

On the other hand, Schanuel’s conjecture also implies a strengthening of Baker’s theorem about the algebraic independence of \( (\mathbb{Q} \text{ linearly independent}) \log \alpha_1, \ldots, \log \alpha_4 \) be four non-zero logarithms of algebraic numbers such that \( \log \alpha_1 / \log \alpha_2, \log \alpha_1 / \log \alpha_3 \notin \mathbb{Q} \). Is it always true that \( \log \alpha_1 \log \alpha_4 - \log \alpha_2 \log \alpha_3 \neq 0 \)? An affirmative answer to this is equivalent to the so called Four Exponentials Conjecture, which is open, but various relaxations of it has been proved, e.g. there is a Five Exponentials Theorem.

1.7. Lower bounds for linear forms in logarithms. Baker’s theorem states that linear combinations of logarithms do not vanish under suitable conditions. The proof, in fact, yields even a lower bound for their absolute value. Since such bounds are very useful in applications, this subject has been revisited by many authors and Baker’s original estimates have been refined.

We are now going to state two sample results in this theory. For the state of the art, we refer the reader to the literature, for which a good starting point is the recent book of Bugeaud [9].

Before we state the results, we make some preliminary remarks. Let \( a_1, \ldots, a_n \in \mathbb{Z}_{>0} \) and \( b_1, \ldots, b_n \in \mathbb{Z} \). We are interested in lower bounds for the linear form in logarithms

\[
|b_1 \log a_1 + \ldots + b_n \log a_n|.
\]

This is closely related to the quantity

\[
|a_1^{b_1} \ldots a_n^{b_n} - 1|.
\]

By expanding \( \exp \) around 0, it is easy to see that when (10) is small, then (11) is also small, and they are within a constant factor. In this situation, the converse is also true. However, later we will consider arbitrary algebraic numbers in place of the \( a_j \), which may not be positive. Then the converse is not necessarily true, because we can only conclude that (10) is close to \( 2\pi i k \) for some \( k \in \mathbb{Z} \) when (11) is small. We will need to pay attention to this issue.

We observe that (11) is a rational number with denominator at most \( a_1^{b_1} \ldots a_n^{b_n} \), and this implies the trivial bound

\[
|b_1 \log a_1 + \ldots + b_n \log a_n| \geq \frac{1}{2} \exp(-(\log a_1 + \ldots + \log a_n)B),
\]
where $B = \max(|b_1|, \ldots, |b_n|)$ provided the left hand side does not vanish. The reason for writing the right hand side in this eccentric way is to make comparisons with the below results easier.

First we state a general result for inhomogeneous forms.

**Theorem 17.** Let $n \in \mathbb{Z}_{\geq 1}$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{Q}_{\neq 0}$ and let $\log \alpha_j$ be any choice of the logarithm of $\alpha_j$. Let $\beta_0, \ldots, \beta_n \in \mathbb{Q}$ and let

$$\Lambda = \beta_0 + \beta_1 \log \alpha_1 + \ldots + \beta_n \log \alpha_n.$$  

Let

$$A_j := \max(H(f_{\alpha_j}), \exp(|\log \alpha_j|), 2)$$

for $j = 1, \ldots, n$ and let

$$B := \max(H(f_{\beta_0}), \ldots, H(f_{\beta_n}), 2).$$

Then there exists an effective constant $C$ depending only on $n$ and the degree of $\mathbb{Q}(\alpha_1, \ldots, \alpha_n, \beta_0, \ldots, \beta_n)$ such that the following holds. If $\Lambda \neq 0$, then

$$|\Lambda| > \exp(-C \log A_1 \cdots \log A_n \log B).$$

Next we state a refined bound for homogeneous forms (i.e. $\beta_0 = 0$) with integral coefficients.

**Theorem 18.** Let $n \in \mathbb{Z}_{\geq 1}$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{Q}_{\neq 0}$ and let $\log \alpha_j$ be any choice of the logarithm of $\alpha_j$. Let $b_1, \ldots, b_n \in \mathbb{Z}$ and let

$$\Lambda = b_1 \log \alpha_1 + \ldots + b_n \log \alpha_n.$$  

Let

$$A_j := \max(H(f_{\alpha_j}), \exp(|\log \alpha_j|), 2)$$

for $j = 1, \ldots, n$ and let

$$B^* := \max(|b_1| \cdot \frac{\log A_1}{\log A_n}, \ldots, |b_n| \cdot \frac{\log A_n}{\log A_n}, 2).$$

Then there exists an effective constant $C$ depending only on $n$ and the degree of $\mathbb{Q}(\alpha_1, \ldots, \alpha_n)$ such that the following holds. If $\Lambda \neq 0$, then

$$|\Lambda| > \exp(-C \log A_1 \cdots \log A_n \log B^*).$$

Some remarks are in order.

- Thanks to Baker’s theorem we could dispose of with the condition $\Lambda \neq 0$ if we assumed that $\log \alpha_j$ are $\mathbb{Q}$ linearly independent. However, it is useful in applications not to restrict ourselves to that setting.
- Theorems 17 and 18 dramatically improves the dependence on the parameter $B$ compared to the trivial bound. This comes at the expense of the dependence on the $A_j$’s. It is expected by some that this trade off is not necessary, and a bound of the form $|\Lambda| > \exp(-C \max(\log A_1, \ldots, \log A_n, \log B))$ may hold. For a more precise prediction, see [23, Conjecture 14.25].
• The improvement in Theorem 18 may seem very minor. The difference between the quantities $B$ and $B^*$ is of any significance only when $A_n$ is much larger than the other $A_j$. However, when $\log A_n$ is very large, the quantity $B^*$ offers a small improvement, which is sometimes important in applications.

• In the special case when the $\alpha_j$ are close to 1, a further improvement over the bounds given in the previous results has been achieved. This is useful in some applications, but we will not discuss these, and we refer to the literature for the details.

• Explicit values of the constant $C$ in the above results have been determined. These are quite large but reasonable, e.g. running into the billions or trillions in Theorem 18 if $n$ and the degree of $Q(\alpha_1, \ldots, \alpha_n)$ is small. However, in the special case $n = 2$, more refined estimates allow for much smaller constants, as small as even 25.2 when everything is rational. This often enables the complete resolution of Diophantine problems (i.e. listing all solutions of some Diophantine equations).

1.8. Digit expansions revisited. As a first application, we deduce a variant of Proposition 9.

**Proposition 19.** There is an effective absolute constant $C$ such that the following holds. For all $p, q, k, m \in \mathbb{Z}_{>0}$, we have

$$|p2^k - q3^m| \geq \frac{\max(2^k, 3^m)}{\max(p, q, 3)^{-C \log \max(k, m) / \log \max(p, q, 3)}}$$

or $p2^k = q3^m$.

Before the proof we make some comments.

• If we apply the weaker bounds in Theorem 17, we get the same result with the exponent $-C \log \max(k, m)$ in place of $-C \log \max(k, m) / \log \max(p, q, 3))$.

• In addition to being effective, this result improves on Proposition 9 when $\max(p, q) < \max(2^k, 3^m)\delta$ for some $\delta > 0$ depending on $\varepsilon$ and the constant $C$.

• In particular, we get

$$|2^k - 3^m| \geq \frac{\max(2^k, 3^m)}{\max(k, m)^{-C}}$$

for some $C$ instead of

$$|2^k - 3^m| \geq \varepsilon \max(2^k, 3^m)^{1-\varepsilon}.$$

• On the other hand, the scope of the $S$-adic subspace theorem is wider. For example, it is easy to give lower bounds for quantities like

$$p_12^{k_1} + p_23^{k_2} + p_35^{k_3}$$
using the method of proof of Proposition 9. However, this seems beyond reach using linear forms in logarithms.

**Proof.** We write $A = \max(p, q, 3)$ and assume for simplicity that $3^m > 2^k$. Suppose further that $p^{2^k} \neq q^{3^m}$. We apply Theorem 18 for

$$
\Lambda = 1 \cdot \log(p/q) + k \log 2 - m \log 3
$$

and get

$$
| \exp \Lambda - 1 | = \left| \frac{p^{2^k}3^{-m} - 1}{q} \right| \geq \exp \left( - C \log A \log \frac{\max(k, m)}{\log A} \right).
$$

Here we do not need to worry about exponentiating $\Lambda$, because all the $\alpha_j$ are positive.

Now we multiply the left hand side by $q3^m$ and the right by $3^m$ and get

$$
| p^{2^k} - q^{3^m} | \geq \frac{3^m}{A^{C \log(\max(k, m)/\log A)}}.
$$

This is precisely what we wanted to prove, since $3^m = \max(2^k, 3^m)$ and $A = \max(p, q, 3)$. □

Recall that for integers $a$ and $b \geq 2$, we write $N(a, b)$ for the number of non-zero digits in the base $b$ expansion of $a$. Using the above result, we can get an improvement of Theorem 10.

**Theorem 20** (Stewart). *There is an effective absolute constant $C$ such that

$$
N(a, 2) + N(a, 3) \geq \frac{\log \log a}{\log \log \log a + C} - 1
$$

for all $a \in \mathbb{Z}_{>2}$.*

Again, here 2 and 3 may be replaced by a pair of multiplicatively independent integers $\geq 2$.

**Proof.** By setting $C$ sufficiently large, we can ensure that the theorem is vacuous for all $a$ smaller than any fixed constant. We may therefore assume without loss of generality that $a$ is suitably large as required by what follows.

Suppose to the contrary that the claim does not hold for some $a \in \mathbb{Z}_{>0}$ with a suitable choice of $C$, which will be made later.

We recall some notation from the proof of Theorem 10. We write $\alpha_k \alpha_{k-1} \ldots \alpha_0$ and $\beta_m \beta_{m-1} \ldots \beta_0$ for the base 2 and base 3 expansions of $a$ respectively. For an interval $I \subset [0, 1]$, we write

$$
\alpha_I := \sum_{j: \log_a(2^j) \in I} \alpha_j 2^j - \min(i: \log_a(2^i) \in I).
$$

In other words, $\alpha_I$ is the integer whose base 2 expansion is the sequence of those $\alpha_j$ for which $2^j = a^t$ for some $t \in I$. We define $\beta_I$ using the base 3 expansion of $a$ in a similar manner.
By Proposition 19, there is some $C_0$ such that

$$|p 2^{k'q} - q 3^{m'}| \geq \frac{\max(2^{k'},3^{m'})}{\max(p,q,2)C_0 \log k}$$

for all $p, q, k', m' \in \mathbb{Z}_{>0}$ with $\max(k', m') \leq k$. Write $K = C_0 \log k + 2$.

By our assumption on $a$, there is some $j \in \{1, \ldots, \log \log \log \log a + C\}$ such that

$$\alpha \left(1 - \frac{K_j + 1}{\log a}, 1 - \frac{K_j}{\log a}\right) = 0.$$

By setting $C$ sufficiently large, which can be done independently of $a$, we can ensure that all of these intervals are contained in $[0,1]$. Also by requiring that $a$ is large enough, we can ensure that each of the numbers in the above equation contain at least 1 digit for all $j$ in the range we consider.

This means that we have $a = p 2^{k_j} + e_1$, where $p = \alpha_{[1-K^j+1/\log a,1-K^j/\log a]}$, $e_1 = \alpha_{[0,1-K^j+1/\log a]}$ and $k_j$ is the smallest integer so that $2^{k_j} \geq a^{1-K^j/\log a}$.

We note that

$$p \leq a/2^{k_j} \leq a^{K^j/\log a}$$

$$e_1 \leq 2a^{1-K^j+1/\log a}.$$

The most important thing for us, as we will see below, is that $e_1 < 2^{k_j}/p^{K^j-2}$. In a similar manner, we can write $a = q 3^{m_j} + e_2$ with

$$q \leq a^{K^j/\log a}$$

$$e_2 \leq 3a^{1-K^j+1/\log a}.$$

Therefore, we have

$$|p 2^{k_j} - q 3^{m_j}| = |e_1 - e_2| \leq \frac{2^{k_j}}{\max(p,q)K^j-2},$$

which contradicts (12). \qed

1.9. **Effective Diophantine approximation.** In this section, we discuss another application of linear forms of logarithms due to Feldman giving an effective improvement of Liouville’s inequality.

**Theorem 21** (Feldman). Let $\alpha$ be an algebraic number of degree $d \geq 3$. Then there are effective constants $c = c(\alpha)$ and $\varepsilon = \varepsilon(\alpha)$ such that

$$\left| \alpha - \frac{p}{q} \right| > \frac{c}{q^{d-\varepsilon}}.$$

For the proof, we need some auxiliary results, which will be more convenient to prove later when we discuss heights in detail.

Let $K$ be a number field (a finite degree extension of $\mathbb{Q}$) and let $\alpha \in K$. Recall that the norm $N_{K|\mathbb{Q}}(\alpha)$ of $\alpha$ in the field extension $K|\mathbb{Q}$
is defined as the determinant of the linear transformation \( x \mapsto \alpha \cdot x \) on \( K \) considered as a vector space over \( Q \). Writing \( \alpha_1, \ldots, \alpha_d \) for the Galois conjugates of \( \alpha \) including itself (that is the roots of its minimal polynomial), we have
\[
N_{K|Q}(\alpha) = (\alpha_1 \cdots \alpha_d)^{[K:Q(\alpha)]}.
\]

**Proposition 22.** For every number field \( K \), there is an integer \( r \in Z \geq 0 \), \( u_1, \ldots, u_r \in K \) and an effective constant \( C = C(K) \) such that the following holds. For every algebraic integer \( \alpha \in K \), there is another algebraic integer \( \tilde{\alpha} \in K \) and integers \( b_1, \ldots, b_r \in Z \) such that
\[
\alpha = \tilde{\alpha}u_1^{b_1} \cdots u_r^{b_r}
\]
\[
H(f_{\tilde{\alpha}}) \leq C|N_{K|Q}(\alpha)|
\]
\[
|b_j| \leq C\log \max(H(f_{\alpha}), 2).
\]

The algebraic numbers \( u_1, \ldots, u_r \in K \) form, in fact, a fundamental system of units, and the proposition can be easily deduced from Dirichlet’s unit theorem, but we do not need to know this at this stage.

We will also need to control the heights of products.

**Proposition 23.** For all \( d \in Z \geq 1 \), there is an effective constant \( C = C(d) \) such that for all algebraic numbers \( \alpha_1, \alpha_2 \) of degree at most \( d \), we have
\[
H(f_{\alpha_1 \alpha_2}) \leq CH(f_{\alpha_1}|Q(\alpha_1):Q(\alpha_2)|H(f_{\alpha_2}|Q(\alpha_2):Q(\alpha_1))\]

This result can be deduced from a variant for another notion of height, denoted by \( H(\cdot) \). The inequality for that simply reads \( H(\alpha_1 \alpha_2) \leq H(\alpha_1)H(\alpha_2) \). We will discuss this in the next section.

Before giving the proof, we discuss its strategy. We aim to find two algebraic numbers of the form \( \alpha_1 \alpha_2^{b_2} \cdots \alpha_n^{b_n} \) that are close to each other and then we will use an argument similar to the proof of Proposition 19 in the previous section. Thanks to Proposition 22, we know that algebraic integers of small norm are of the required form.

Without loss of generality, we can assume that \( \alpha \) is an algebraic integer. Let \( P \in Z[x] \) be the minimal polynomial of \( \alpha \). Write
\[
P(x) = (x - \alpha_1) \cdots (x - \alpha_d)
\]
with \( \alpha_1 = \alpha \). Let \( p/q \in Q \) such that it is closest to \( \alpha_1 \) among the roots.

As we seen in Lemma 5, the absolute value of
\[
(p - \alpha_1q) \cdots (p - \alpha_dq)
\]
is bounded above and below by a constant multiple of \( q^d|\alpha_1 - p/q| \).

Now we assume that \( (13) \leq q^\varepsilon \) for a suitable \( \varepsilon > 0 \) and aim to show that this can hold only if \( p \) and \( q \) are bounded by an effective constant depending only on \( \alpha \).

Since \( \alpha_1, \ldots, \alpha_d \) are algebraic integers, we have that \( N_{Q(\alpha)|Q}(p - \alpha_jq) \) is an integer dividing \( (13)^j \) for each \( j \). This means that \( p - \alpha_jq \) is of the
required form, and we could run our argument if we could find two of these that are close to each other. Unfortunately, this is not the case, but we know that \(|p - \alpha_1q|\) is small (if \(|\alpha_1 - p/q|\) is small) and hence \(p - \alpha jq\) is close to \((\alpha_1 - \alpha jq)\) for each \(j\). Therefore, \((\alpha_1 - \alpha_2)(p - \alpha_3q)\) and \((\alpha_1 - \alpha_3)(p - \alpha_2q)\) are close to each other and both of these numbers can be written in the required form.

We carry out the proof motivated by the above observations.

**Proof of Theorem 2.1.** We assume without loss of generality that \(\alpha\) is an algebraic integer. We fix a small \(\varepsilon > 0\). Let \(p/q \in \mathbb{Q}\) be such that
\[
|\alpha - p/q| < q^{-d+\varepsilon}.
\]

Let
\[
P(x) = (x - \alpha_1) \cdots (x - \alpha_d)
\]
be the minimal polynomial of \(\alpha\), and let \(\alpha_1 = \alpha\).

We compute
\[
\left| 1 - \frac{(\alpha_1 - \alpha_3)(p - \alpha_2q)}{(\alpha_1 - \alpha_2)(p - \alpha_3q)} \right| = \left| 1 - \frac{(\alpha_1 - \alpha_3)(p - \alpha_1q + (\alpha_1 - \alpha_2)q)}{(\alpha_1 - \alpha_2)(p - \alpha_1q + (\alpha_1 - \alpha_3)q)} \right|
\]
\[
\leq \left| 1 - \frac{(\alpha_1 - \alpha_3)(\alpha_1 - \alpha_2)q}{(\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)q} \right| + C q^{-d+\varepsilon}
\]
\[
= C q^{-d+\varepsilon}
\]
for a constant \(C\) depending only on \(\alpha\). In the rest of the proof \(C\) will denote a constant that depends only on \(\alpha\) whose value may change between occurrences.

By Lemma 5, we have
\[
| (p - \alpha_1q) \cdots (p - \alpha_dq) | \leq C q^\varepsilon.
\]

Since \(p - \alpha_2q\) is an algebraic integer, its norm (in the field extension \(\mathbb{Q}(\alpha_2)/\mathbb{Q}\)) is an integer that divides \((p - \alpha_1q)^d \cdots (p - \alpha_dq)^d\). By Proposition 22, we can write
\[
p - \alpha_2q = \tilde{\alpha}_2u_1^{b_1} \cdots u_r^{b_r},
\]
where \(r \in \mathbb{Z}_{\geq 0}, u_1, \ldots, u_r\) are elements in \(\mathbb{Q}(\alpha_2)\) independent of \(p\) and \(q\), and \(\tilde{\alpha}_2 \in \mathbb{Q}(\alpha_2), b_1, \ldots, b_r \in \mathbb{Z}\) satisfy
\[
H(f_{\tilde{\alpha}_2}) \leq C q^{d_2}, \quad \max(|b_1|, \ldots, |b_r|) \leq C \log q.
\]
Here we used that \(H(f_{p - \alpha_2q}) \leq C q^d\), which can be seen by calculating the minimal polynomial of \(p - \alpha_2q\) in terms of \(P\).

Similarly, we can write
\[
p - \alpha_2q = \tilde{\alpha}_3w_1^{e_1} \cdots w_r^{e_r},
\]
where \(r \in \mathbb{Z}_{\geq 0}, w_1, \ldots, w_r\) are elements in \(\mathbb{Q}(\alpha_3)\) independent of \(p\) and \(q\), and \(\tilde{\alpha}_3 \in \mathbb{Q}(\alpha_3), e_1, \ldots, e_r \in \mathbb{Z}\) satisfy
\[
H(f_{\tilde{\alpha}_3}) \leq C q^{d_3}, \quad \max(|e_1|, \ldots, |e_r|) \leq C \log q.
\]
Now we write

$$\alpha^* = \frac{(\alpha_1 - \alpha_3)\bar{\alpha}_2}{(\alpha_1 - \alpha_2)\bar{\alpha}_3}$$

and note that

$$H(f_{\alpha^*}) \leq Cq^{Ce}$$

by Proposition 23. Plugging in to our estimate at the beginning of the proof, we get

$$\left|1 - \alpha^*u_1^{b_1} \cdots u_r^{b_r} w_1^{-e_1} \cdots w_r^{-e_r}\right| \leq Cq^{-d+\varepsilon}.$$ 

We take logarithm of the product on the left choosing the principal branch of log for $\alpha^*$ and for each $u_j$ and $w_j$, that is, the branch with $|\text{Im}(\log(\cdot))| \leq \pi$. We get

$$|\log \alpha^* + b_1 \log u_1 + \cdots + b_r \log u_r - e_1 \log w_1 - \cdots - e_r \log w_r + 2k \log(-1)| \leq Cq^{-d+\varepsilon},$$

where $k$ is an integer satisfying

$$|k| \leq C \max(|b_1|, \ldots, |b_r|, |e_1|, \ldots, |e_r|) \leq C \log q.$$ 

We apply Theorem 18 with $\alpha^*$ in the role of $\alpha_n$ and the $u_j$, $w_j$ and $-1$ in the role of the other $\alpha_i$. We have $A_i \leq C$ for all $i \neq n$, $A_n \leq Cq^{Ce}$ and

$$B \leq \max\left(\frac{C \log q}{\log A_n}, 2\right) \leq C \varepsilon^{-1}.$$ 

Theorem 18 yields

$$|\log \alpha^* + b_1 \log u_1 + \cdots + b_r \log u_r - e_1 \log w_1 - \cdots - e_r \log w_r + 2k \log(-1)| \geq \exp(-C \log A_n \log \varepsilon^{-1}) \geq C^{-1} q^{-C \varepsilon \log \varepsilon^{-1}}.$$ 

Choosing $\varepsilon$ sufficiently small so that $C \varepsilon \log \varepsilon^{-1} < d - \varepsilon$, we can conclude that $|\alpha - p/q| < q^{-d+\varepsilon}$ may hold only if $q$ is bounded by an effective constant depending only on $\alpha$. This proves the theorem. $\square$

If we apply the weaker bound in Theorem 17 instead of Theorem 18, we could get the slightly weaker conclusion

$$|\alpha - p/q| \geq cq^{-\varepsilon/\log \log q}.$$ 

2. Heights

We have already encountered the quantity $H(f_\alpha)$, the maximal absolute value of the coefficients of the minimal polynomial of $\alpha \in \overline{Q}$, which we called the naive height of $\alpha$. This is a very natural quantity to measure the “complexity” of an algebraic number. However, it is a bit cumbersome to use $H(f_\alpha)$ in calculations, as the coefficients of the minimal polynomial is difficult to control when we perform operations on algebraic numbers. For this reason, we introduce another very natural quantity to measure complexity and discuss its properties in this section.
We follow the notation and exposition of Masser in [16, Chapter 14]. There are many alternative references, where this subject is discussed, see for example Bombieri and Gubler [7, Chapter 1], whose notation differs from ours in certain places.

2.1. Mahler measure. Let

\[ P(x) = a_d(x - \alpha_1) \cdots (x - \alpha_d) = a_dx^d + \ldots + a_0 \in \mathbb{C}[x] \]

be a polynomial. The **Mahler measure** of \( P \) is defined by

\[ M(P) = |a_d| \prod_{j=1}^{d} \max(1, |\alpha_j|). \]

When \( \alpha \) and \( \beta \) are algebraic integers, it is easy to bound the Mahler measure of the minimal polynomials of \( \alpha + \beta \) and \( \alpha\beta \) in terms of the Mahler measures of the minimal polynomials of \( \alpha \) and \( \beta \). Indeed, we write \( \alpha_1, \ldots, \alpha_{d_1} \) for the Galois conjugates of \( \alpha \) including itself and \( \beta_1, \ldots, \beta_{d_2} \) for the Galois conjugates of \( \beta \) again including itself. For simplicity, we assume that \( [\mathbb{Q}(\alpha \beta) : \mathbb{Q}] = [\mathbb{Q}(\alpha + \beta) : \mathbb{Q}] = d_1d_2 \). Then we can write

\[
M(f_{\alpha \beta}) = \prod_{i=1}^{d_1} \prod_{j=1}^{d_2} \max(1, |\alpha_i \beta_j|)
\]

\[
\leq \prod_{i=1}^{d_1} \prod_{j=1}^{d_2} \max(1, |\alpha_i|) \max(1, |\beta_j|)
\]

\[
= M(f_\alpha)^{d_2} M(f_\beta)^{d_1}
\]

and

\[
M(f_{\alpha + \beta}) = \prod_{i=1}^{d_1} \prod_{j=1}^{d_2} \max(1, |\alpha_i + \beta_j|)
\]

\[
\leq \prod_{i=1}^{d_1} \prod_{j=1}^{d_2} 2 \max(1, |\alpha_i|) \max(1, |\beta_j|)
\]

\[
\leq 2^{d_1d_2} M(f_\alpha)^{d_2} M(f_\beta)^{d_1}.
\]

When \( \alpha \) or \( \beta \) are not algebraic integers, then we need to deal with the leading coefficients, which is not completely straightforward. To address this issue, we will define the height \( H(\alpha) \) of an algebraic number in a different way in the next section. However, we will show that \( H(\alpha) = M(f_\alpha)^{1/d} \), where \( d \) is the degree of \( \alpha \).

Before we do this, we explore the relation between Mahler measures and naive heights. To this end, it is useful to note the formula

\[
(14) \quad \log M(P) = \int_0^1 \log |P(e^{2\pi it})|dt.
\]
for the Mahler measure, which follows easily from Jensen’s formula.

**Theorem 24.** Let \( P \in \mathbb{C}[x] \) of degree \( d \). Then
\[
2^{-d} H(P) \leq M(P) \leq (d + 1) H(P).
\]

**Proof.** To prove the upper bound, we use (14). We can write
\[
\log M(P) \leq \int_0^1 \log |P(e^{2\pi it})| dt
\]
\[
\leq \int_0^1 \log((d + 1)H(P)) dt
\]
\[
= \log((d + 1)H(P)),
\]
and the claim follows by exponentiation.

To prove the lower bound, we can estimate the coefficients by
\[
\frac{|a_k|}{|a_d|} \leq \prod_{\{j_1, \ldots, j_k\} \subset \{1, \ldots, d\}} |\alpha_{j_1}| \cdots |\alpha_{j_k}| \leq 2^d \prod_{j=1}^d \max(1, |\alpha_j|).
\]
Multiplying by \(|a_d|\) and dividing by \(2^d\) both sides and taking the maximum over \(k\), we get \(2^{-d}H(P) \leq M(P)\) as required. \(\square\)

A closer inspection of the proof reveals that \(H(P)\) could be replaced by the \(l^1\)-norm of the coefficients in the lower bound, and \((d + 1)H(P)\) could be replaced by the same in the upper bound. In fact, the upper bound could be replaced even by the \(l^2\)-norm if we exploit the orthogonality of the functions \(t \mapsto e^{2\pi ikt}\).

2.2. **Heights.** As we mentioned above, we would like a better way of dealing with the leading coefficient of the minimal polynomial. It turns out that this can be done in a nice way using absolute values coming from not only the embeddings of \(\alpha\) into \(\mathbb{C}\) but also into \(p\)-adic fields.

We introduce the required notation. Let \(K\) be a number field, and let \(\mathcal{O}_K\) be its ring of integers. We denote by \(\Sigma_K\) the places of \(K\), which comprises all prime ideals of \(\mathcal{O}_K\) and all embeddings of \(K\) into \(\mathbb{C}\) (some of these may map \(K\) to \(\mathbb{R}\)). We write \(\Sigma_{K,f}\) for the finite places (i.e. the prime ideals) and \(\Sigma_{K,\infty}\) for the infinite places (i.e. the embeddings to \(\mathbb{C}\)). For each place \(v \in \Sigma_K\), we introduce an absolute value (or valuation) denoted by \(|\cdot|_v\) on \(K\). If \(v\) is finite, that is a prime ideal, then we define
\[
|\alpha|_v = N(v)^{-k}
\]
for \(\alpha \in \mathcal{O}_K\), where \(N(v)\) is the norm of \(v\) (as an ideal), and \(k\) is the largest integer such that \(v^k|\alpha\mathcal{O}_K\). We extend this to \(K\) multiplicatively. If \(v\) is infinite, that is an embedding \(K \to \mathbb{C}\), we define
\[
|\alpha|_v = |v(\alpha)|,
\]
where \(|\cdot|\) is the standard absolute value on \(\mathbb{C}\).
We digress to discuss an important aspect of our terminology. When we say that \( \Sigma_K \) contains all embeddings into \( \mathbb{C} \), we really mean all of them and not just one from each pair of complex conjugates. This terminology, following Masser [16, Chapter 14], is slightly eccentric in that we treat the finite and infinite places differently. (We only included one from each collection of equivalent absolute values coming from \( p \)-adic embeddings.) An alternative (and perhaps more common) approach would be to include only one embedding from each pair of complex conjugates. Doing this requires changing the normalization of \( |\cdot|_v \). In the book of Bombieri and Gubler [7] this is done by taking our definition to the power \( \frac{1}{[K: \mathbb{Q}]} \) if \( v \) is a prime ideal or an embedding with real image and taking it to the power \( \frac{2}{[K: \mathbb{Q}]} \) if \( v \) is an embedding with complex image. The disadvantage of this approach is that \( |\cdot|_v \) is no longer the restriction of the standard absolute value for infinite places.

We made the choice of our normalization so that the following product formula holds

**Lemma 25.** For all \( \alpha \in K \) we have

\[
\prod_v |\alpha|_v = 1.
\]

When we write \( \prod_v \) as above without specifying the range of \( v \) we mean the product over all places.

**Proof.** If \( \alpha \in \mathcal{O}_K \), this follows easily by

\[
\prod_{v \in \Sigma_{K,f}} |\alpha|_v^{-1} = N(\alpha \mathcal{O}_K) = |N_{K|\mathbb{Q}}(\alpha)| = \prod_{v \in \Sigma_{K,\infty}} v(\alpha).
\]

For general \( \alpha \) it follows by multiplicativity of the absolute values. \( \square \)

We define the **absolute height** or **Weil height** of \( \alpha \in K \) by

\[
H(\alpha) = \prod_v \max(1, |\alpha|_v)^{1/[K: \mathbb{Q}]}
\]

In the literature, the notation \( h(\alpha) = \log H(\alpha) \) is also commonly used and it is called the logarithmic (absolute/Weil) height \( \alpha \).

**Proposition 26.** The above definition of \( H(\alpha) \) is independent of the choice of the number field \( K \) containing \( \alpha \).

**Proof.** Let \( L|K \) be a field extension. We need to show that

\[
\prod_{v \in \Sigma_K} \max(1, |\alpha|_v)^{1/[K: \mathbb{Q}]} = \prod_{v \in \Sigma_L} \max(1, |\alpha|_v)^{1/[L: \mathbb{Q}]}
\]

It is a fact of Galois theory that for each embedding \( v : K \to \mathbb{C} \), there are \([L : K]\) many embeddings \( w : L \to \mathbb{C} \) such that \( w|_K = v \). This implies that infinite places contribute equally to both sides of (15).
To show the same for finite places, we make the following considerations. Let $P$ be a prime ideal of $\mathcal{O}_K$. Then $P\mathcal{O}_L$ is an ideal of $\mathcal{O}_L$ and its norm is $N(P\mathcal{O}_L) = N(P)^{[L:K]}$. We can factorize this into a product of prime ideals as

$$P\mathcal{O}_L = P_1^{e_1} \cdots P_k^{e_k}.$$ 

We note that no prime ideal of $L$ can occur in the factorizations of two different prime ideals of $K$.

Now let $\alpha \in \mathcal{O}_K$. If $P$ occurs with exponent $m$ in the factorization of $\alpha \mathcal{O}_K$, then $P_j$ occurs with exponent $e_j m$ in $\alpha \mathcal{O}_L$. This, and

$$N(P_1)^{e_1} \cdots N(P_k)^{e_k} = N(P\mathcal{O}_L) = N(P)^{[L:K]}$$

implies

$$|\alpha|_{P_1} \cdots |\alpha|_{P_k} = |\alpha|^{[L:K]}_P,$$

where the absolute values on the left correspond to places of $L$ and the absolute value on the right corresponds to a place of $K$. We can extend this identity to all $\alpha \in K$ by multiplicativity. Furthermore, the same considerations imply that for any $j$, $|\alpha|_{P_j} > 1$ is equivalent to $|\alpha|_P > 1$. This shows that the finite places also contribute equally to both sides of (15).

\[\Box\]

**Proposition 27.** We have

$$H(\alpha) = M(f_\alpha)^{1/[Q(\alpha):Q]}.$$

In light of this result and our estimates for the Mahler measure in terms of the naive height, we have

$$2^{-d}H(f_\alpha) \leq H(\alpha)^{[Q(\alpha):Q]} \leq (d + 1)H(f_\alpha).$$

**Proof.** Comparing the definitions, it is enough to show that

$$\prod_{v \in \Sigma_{K.f}} \max(1, |\alpha|_v) = |a_d|^{[K:Q(\alpha)]},$$

where $a_d$ is the leading coefficient of the minimal polynomial of $\alpha$. Strictly speaking, the sufficiency of this can be seen from the definitions for $K = Q(\alpha)$, but the proof of the previous proposition shows that we can take any number field $K$ containing $\alpha$. We take $K$ to be a field that contains all Galois conjugates $\alpha_1, \ldots, \alpha_d$ of $\alpha$.

The proof relies on the following version of Gauss's lemma. For a polynomial $Q \in K[x]$ and a place $v$, we write $|Q|_v$ for the maximum of the $|\cdot|_v$ absolute values of the coefficients of $Q$. Gauss's lemma states that for any $Q_1, Q_2 \in K[x]$ and $v \in \Sigma_{K.f}$, we have

$$|Q_1 Q_2|_v = |Q_1|_v \cdot |Q_2|_v.$$ 

Now we apply this to the factorization of the minimal polynomial of $\alpha$. By definition of the minimal polynomial, we have

$$|a_d(x - \alpha_1) \cdots (x - \alpha_d)|_v = 1.$$
Therefore,
\[
\prod_{v \in \Sigma_{K,f}} |a_d|_v \cdot \prod_{j=1}^d \prod_{v \in \Sigma_{K,f}} \max(1, |\alpha_j|_v) = 1.
\]
Now we use that each \(\alpha_j\) can be mapped to \(\alpha\) by an automorphism of \(K\) and the product formula for \(a_d\). We get
\[
\left( \prod_{v \in \Sigma_{K,f}} \max(1, |\alpha|_v) \right)^d = \prod_{v \in \Sigma_{K,\infty}} |a_d|_v.
\]
This implies (16), because \(|a_d|_v = |a_d|\) for each infinite place. \(\square\)

2.3. Calculating with heights. Now we turn our attention to estimates controlling heights when we perform operations on elements of \(K\).

**Lemma 28.** For all \(\alpha \in K\) and \(k \in \mathbb{Z}\) we have
\[
H(\alpha^k) = H(\alpha)^{|k|}.
\]

**Proof.** This is an immediate consequence of the definition of \(H(\alpha)\) for \(k > 0\), and it follows from the product formula for \(k = -1\). \(\square\)

For a polynomial \(P\) with complex coefficients in possibly several variables, we write \(\mathcal{L}(P)\) for the sum of the absolute values of its coefficients.

**Proposition 29.** Let \(k \in \mathbb{Z}_{\geq 1}\), \(d_1, \ldots, d_k \in \mathbb{Z}_{\geq 0}\), and let \(P, Q \in \mathbb{Z}[x_1, \ldots, x_k]\) be two polynomials that are of degree at most \(d_j\) in the variable \(x_j\) for each \(j\). Let \(\alpha_1, \ldots, \alpha_k \in \overline{Q}\) Then
\[
H\left( \frac{P(\alpha_1, \ldots, \alpha_k)}{Q(\alpha_1, \ldots, \alpha_k)} \right) \leq \max(\mathcal{L}(P), \mathcal{L}(Q)) \prod_{j=1}^k H(\alpha_j)^{d_j}
\]

Before giving the proof, we note the following two special cases of interest
\[
H(\alpha_1 \alpha_2) \leq H(\alpha_1)H(\alpha_2), \quad H(\alpha_1 + \alpha_2) \leq 2H(\alpha_1)H(\alpha_2).
\]

In light of the relationship between \(H(f_\alpha)\) and \(H(\alpha)\), which we established above, this implies Proposition 23.

**Proof.** Let \(K\) be a number field containing all \(\alpha_j\). For a finite place \(v\), the ultrametric property implies that \(|P(\alpha_1, \ldots, \alpha_k)|_v\) can be estimated by the maximal absolute value of a monomial. Since the coefficients of \(P\) are in \(\mathbb{Z}\) whose \(|\cdot|_v\) absolute value are at most 1, we can disregard them in our calculation. We can thus write
\[
|P(\alpha_1, \ldots, \alpha_k)|_v \leq \max_{m_j=0, \ldots, d_j} \prod_{j=1}^k |\alpha_j^{m_j}|_v = \prod_{j=1}^k \max(1, |\alpha_j|_v)^{d_j}.
\]
For an infinite place \( v \), we use the triangle inequality for the monomials and get

\[
|P(\alpha_1, \ldots, \alpha_k)|_v \leq \mathcal{L}(P) \max_{m_j=0, \ldots, d_j} \prod_{j=1}^{k} |\alpha_j^{m_j}|_v
\]

\[
= \mathcal{L}(P) \prod_{j=1}^{k} \max(1, |\alpha_j|_v)^{d_j}.
\]

We can write

\[
H\left(\frac{P(\alpha_1, \ldots, \alpha_k)}{Q(\alpha_1, \ldots, \alpha_k)}\right)_{[K:Q]} = \prod_v \max\left(1, \frac{|P(\alpha_1, \ldots, \alpha_k)|_v}{|Q(\alpha_1, \ldots, \alpha_k)|_v}\right).
\]

Using the product formula for \( Q(\alpha_1, \ldots, \alpha_d) \), we can rewrite this as

\[
H\left(\frac{P(\alpha_1, \ldots, \alpha_k)}{Q(\alpha_1, \ldots, \alpha_k)}\right)_{[K:Q]} = \prod_v \max\left(|Q(\alpha_1, \ldots, \alpha_k)|_v, |P(\alpha_1, \ldots, \alpha_k)|_v\right).
\]

Now we plug in our estimates for the valuations of \( P(\alpha_1, \ldots, \alpha_k) \) and similar bounds for those of \( Q(\alpha_1, \ldots, \alpha_k) \). Since there are \([K : \mathbb{Q}]\) infinite places, we get

\[
H\left(\frac{P(\alpha_1, \ldots, \alpha_k)}{Q(\alpha_1, \ldots, \alpha_k)}\right)_{[K:Q]} \leq \max(\mathcal{L}(P), \mathcal{L}(Q))^{[K:Q]} \prod_{j=1}^{k} \max(1, |\alpha_j|_v)^{d_j}
\]

\[
= \max(\mathcal{L}(P), \mathcal{L}(Q))^{[K:Q]} \prod_{j=1}^{k} H(\alpha_j)^{d_j}[K:Q],
\]

which proves the theorem. \( \square \)

**Lemma 30.** For all \( \alpha \in \overline{\mathbb{Q}} \), we have

\[
H(\alpha)^{-d} \leq |\alpha| \leq H(\alpha)^d.
\]

**Proof.** The upper bound follows directly from the definition of \( H(\alpha) \). The lower bound follows from the upper bound applied for \( \alpha^{-1} \). \( \square \)

2.4. **Units.** The purpose of this section is to prove the following result, which we used earlier in the proof of Feldman’s theorem on the effective improvement of Liouville’s inequality.

**Proposition 31.** For every number field \( K \), there is an integer \( r \in \mathbb{Z}_{\geq 0} \), \( u_1, \ldots, u_r \in K \) and an effective constant \( C = C(K) \) such that the following holds. For every algebraic integer \( \alpha \in K \), there is another algebraic integer \( \tilde{\alpha} \in K \) and integers \( b_1, \ldots, b_r \in \mathbb{Z} \) such that

\[
\alpha = \tilde{\alpha} u_1^{b_1} \cdots u_r^{b_r}
\]

\[
H(f_{\tilde{\alpha}}) \leq C|N_{K|\mathbb{Q}}(\alpha)|
\]

\[
|b_j| \leq C \log \max(H(f_{\tilde{\alpha}}), 2).
\]
The proof of this result relies on Dirichlet’s unit theorem, which we recall now together with the relevant notation. Let $K$ be a number field of degree $d$. We consider the map

$$\Phi : K \times \to R^{\Sigma_{K,\infty}} \equiv R^d, \quad \alpha \mapsto (\log |\alpha|_v)_{v \in \Sigma_{K,\infty}},$$

which is a homomorphism from the multiplicative group of $K$ to the additive group $R^d$. It is clear that $\Phi(K\times)$ is contained in the subspace of points $(x_v) \in R^{\Sigma_{K,\infty}}$ that satisfy $x_v = x_{\overline{v}}$ for each pair of complex conjugate embeddings $v, \overline{v} \in \Sigma_{K,\infty}$. We denote this subspace by $W$. At this point it would be slightly more convenient to use the convention that $\Sigma_{K,\infty}$ contains only one from each pair of complex conjugate embeddings, since then $W$ could be taken to be the whole space.

It is a theorem of Kronecker that $\text{Ker} \Phi$ is the set of roots of unity contained in $K$. It is immediate from the definitions that

$$|N_{K|Q}(\alpha)| = \exp \left( \sum_v (\Phi(\alpha))_v \right)$$

for all $\alpha \in K\times$ and

$$\exp(\|\Phi(\alpha)\|_1/2) \leq H(\alpha)^d \leq \exp(\|\Phi(\alpha)\|_1)$$

for all $\alpha \neq 0 \in \mathcal{O}_K$.

The units of $\mathcal{O}_K$, denoted by $\mathcal{O}_K^\times$, is the set of those elements whose multiplicative inverses are also in $\mathcal{O}_K$. It is immediate from this definition that $\alpha \in \mathcal{O}_K$ is a unit if and only if $|N_{K|Q}(\alpha)| = 1$. We write $W_0$ for the 1-codimensional subspace of $W$ satisfying the equation

$$\sum_v (\Phi(\alpha))_v = 0.$$

Thus $\alpha \in \mathcal{O}_K$ is a unit if and only if $\Phi(\alpha) \in W_0$.

Dirichlet’s unit theorem states that $\Phi(\mathcal{O}_K^\times)$ is a lattice in $W_0$.

**Proof of Proposition 31.** Let $r = \dim W_0$ and let $u_1, \ldots, u_r$ be a fundamental system of units in $\mathcal{O}_K$, that is, $\Phi(u_1), \ldots, \Phi(u_r)$ is a basis of the lattice $\Phi(\mathcal{O}_K^\times)$. Let $e = (1/d, \ldots, 1/d) \in R^{\Sigma_{K,\infty}}$. Then $e \in W \setminus W_0$, and hence $\Phi(u_1), \ldots, \Phi(u_r), e$ is a basis of the vector space $W$.

Let $\alpha \in K$, and let $y_1, \ldots, y_{r+1} \in R$ such that

$$\Phi(\alpha) = y_1 \Phi(u_1) + \ldots + y_r \Phi(u_r) + y_{r+1} e.$$

Then $y_{r+1} = \log |N_{K|Q}(\alpha)|$ and

$$|y_j| \leq C \log H(\alpha)$$

for $j = 1, \ldots, r$, where $C$ is a constant that depends only on $u_1, \ldots, u_r$ so ultimately only on $K$.

We set $b_j = |y_j|$ for $j = 1, \ldots, r$. The required upper bound on $\sum b_j$ follows at once from the estimate on $|y_j|$ and the relationship between $H(f_\alpha)$ and $H(\alpha)$. Furthermore, we set $\tilde{\alpha} = \alpha u_1^{-b_1} \cdots u_r^{-b_r}$. Then
\[ \Phi(\alpha) = x + y_{r+1} \epsilon, \] where \( x \) belongs to the compact set 
\[ \{ z_1 \Phi(u_1) + \ldots + z_r \Phi(u_r) : z_1, \ldots, z_r \in [0, 1]^r \}, \]
which depends only on \( K \). Therefore,
\[ H(\tilde{\alpha})^d \leq \exp(\|\Phi(\tilde{\alpha})\|_1) \leq \exp(C \cdot N_{K:Q}(\alpha)), \]
where \( C \) depends only on \( K \). By the relationship between \( H(\tilde{\alpha}) \) and \( H(f_\alpha) \), this completes the proof. \( \square \)

3. Roth’s theorem – proof

The purpose of this section is to prove Roth’s theorem, which we have already stated in the introduction, and which we recall now.

**Theorem 32 (Roth).** Let \( \alpha \) be an irrational real algebraic number. Then for all \( \epsilon > 0 \), there is a constant \( c = c(\epsilon, \alpha) \) such that
\[ |\alpha - p/q| \geq \frac{c}{q^{2+\epsilon}} \]
for all rational numbers \( p/q \).

A full self-contained exposition can be found in the book of Cassels [11, Chapter VI], from which we borrow heavily.

The strategy of the proof is the following.

1. Assume to the contrary that there are infinitely many \( p/q \in \mathbb{Q} \) with \( |\alpha - p/q| < q^{-2-\epsilon} \) and select a suitable finite subset \( p_1/q_1, \ldots, p_k/q_k \) among them.
2. Find a suitable polynomial \( P(x_1, \ldots, x_k) \in \mathbb{Z}[x_1, \ldots, x_k] \) that vanishes at the point \((\alpha, \ldots, \alpha)\) to high order.
3. Give a lower bound for \( |P(p_1/q_1, \ldots, p_k/q_k)| \).
4. Give an upper bound for \( |P(p_1/q_1, \ldots, p_k/q_k)| \).
5. Realize that the upper bound and the lower bound are in contradiction.

This can be seen as a natural extension of Liouville’s proof, where \( k = 1 \) and \( P \) was chosen to be the minimal polynomial of \( \alpha \). It turns out that it is necessary to consider polynomials in more than one variable to obtain any improvement over Liouville’s theorem. Indeed, if we take any \( P \in \mathbb{Z}[x] \) of degree \( n \), then it can have a zero of multiplicity at most \( n/d \) at \( \alpha \), where \( d = [Q(\alpha) : Q] \). The lower bound on \( |P(p/q)| \) will be \( 1/q^n \) (if we manage to show that \( P(p/q) \neq 0 \)), and the upper bound is of the order \( |\alpha - p/q|^{n/d} \), which gives us no contradiction unless \( |\alpha - p/q| \leq q^{-d} \).

It is reasonable to expect that the lower bound \( 1/q^n \) on \( P(p/q) \) is far from the truth, but nobody seems to be able to do any better.

Therefore, we will consider polynomials in several variables, and also several rational approximations to \( \alpha \). If we substituted the same rational to each variable, then in effect, we would have a polynomial in a single variable, and not much hope for progress. This is unfortunate,
because this is what makes the proof ineffective. Indeed, if there was only one very good rational approximation to \( \alpha \), however good it was, we were not able to reach a contradiction. So our argument is not able to exclude the possibility that one such approximation exists.

Things would be rather different if we were able to find an exceptionally close approximation to \( \alpha \). Then our proof could be used to exclude the possibility that a second one exists and we could get an effective result. The pitfall is, of course, that it is exceedingly unlikely that we would ever find an approximation that is sufficiently close for this purpose, because it probably does not exist.

Now we turn back to our strategy and we begin by analysing its feasibility. We will break up the proof into smaller problems, which we will study in detail in the sequel.

The main task in giving a lower bound on \(|P(p_1/q_1, \ldots, p_k/q_k)|\) will be to show that it does not vanish. This was very easy in the proof of Liouville’s theorem, but it will be the most difficult part of the proof now. In fact, the rest of the proof was already well understood by Siegel, and this was the issue that Roth had to overcome, which earned him the Fields medal. We will return to this point later. Once we show non-vanishing, we easily obtain the lower bound

\[
|P(p_1/q_1, \ldots, p_k/q_k)| \geq q_1^{-n_1} \cdots q_k^{-n_k},
\]

where \( n_j \) denotes the degree of \( P \) in \( x_j \). This follows by observing that the left hand side is a rational number with denominator \( q_1^{n_1} \cdots q_k^{n_k} \).

For the upper bound, we will use the Taylor expansion of \( P \) at \((\alpha, \ldots, \alpha)\). This can be written as

\[
P(p_1/q_1, \ldots, p_k/q_k) = \sum_{j_1, \ldots, j_k} P_{j_1, \ldots, j_k}(\alpha, \ldots, \alpha)(\alpha - p_1/q_1)^{j_1} \cdots (\alpha - p_k/q_k)^{j_k},
\]

where

\[
P_{j_1, \ldots, j_k} = \frac{1}{j_1! \cdots j_k!} \frac{\partial^{j_1 + \cdots + j_k}}{\partial x_1^{j_1} \cdots \partial x_k^{j_k}} P.
\]

For the purposes of this discussion, we will ignore the size of the coefficients \( P_{j_1, \ldots, j_k}(\alpha, \ldots, \alpha) \) and also we will just look at largest non-zero term in (18). Later on we will need to estimate this more precisely, which will also require an estimate on the coefficients of \( P \).

We observe that

\[
|\alpha - p_1/q_1|^{j_1} \cdots |\alpha - p_k/q_k|^{j_k} \leq q_1^{-(2+\varepsilon)j_1} \cdots q_k^{-(2+\varepsilon)j_k} \exp(-(2+\varepsilon)(j_1 \log q_1 + \cdots + j_k \log q_k)).
\]

This motivates us to introduce the notion of the index of a function \( F : \mathbb{R}^k \to \mathbb{R} \) at a point \((\beta_1, \ldots, \beta_k)\) with respect to weights \( w_1, \ldots, w_k \) by

\[
I_F(b_1, \ldots, b_k; w_1, \ldots, w_k) = \min_{j_1, \ldots, j_k} \left\{ j_1 w_1 + \cdots + j_k w_k : F_{j_1, \ldots, j_k}(\beta_1, \ldots, \beta_k) \neq 0 \right\}.
\]
When the weights are clear from the context, we may omit them from our notation. With this notation, we can write our upper bound for (18) in the form

$$\exp(-(2 + \varepsilon) I_P(\alpha, \ldots, \alpha; \log q_1, \ldots, \log q_k)).$$

We stress again that this is not a precise upper bound, and we will return to this point.

Now we consider the problem of choosing the polynomial $P$. Our aim is to choose it in such a way that (17) is larger than (19). To this end, we need to maximize the index of $P$ at $(\alpha, \ldots, \alpha)$ with respect to the weights $\log q_1, \ldots, \log q_k$. We fix some numbers $n_1, \ldots, n_k$ and look for $P$ in the form

$$P = \sum_{j_1=0}^{n_1} \cdots \sum_{j_k=0}^{n_k} a_{j_1, \ldots, j_k} x_1^{j_1} \cdots x_k^{j_k}$$

with $a_{j_1, \ldots, j_k} \in \mathbb{Z}$. For some fixed $i_1, \ldots, i_k$, the equation

$$P_{i_1, \ldots, i_k}(\alpha, \ldots, \alpha) = 0$$

is a linear equation in the variables $a_{j_1, \ldots, j_k}$ with coefficients in $\mathbb{Q}(\alpha)$. If we choose a basis in $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$, this becomes a system of $d = [\mathbb{Q}(\alpha) : \mathbb{Q}]$ linear equations over $\mathbb{Q}$. If we want to find a $P$ such that $I_P(\alpha, \ldots, \alpha) \geq I$ for some $I \in \mathbb{R}_{\geq 0}$, then we need to solve

$$d \cdot \left| \left\{ (i_1, \ldots, i_k) : i_1 \log q_1 + \cdots + i_k \log q_k < I \right\} \right|$$

many equations. We have $(n_1 + 1) \cdots (n_k + 1)$ many variables, so by linear algebra, we can find a non-zero solution $P$ provided

$$(n_1 + 1) \cdots (n_k + 1) > d \cdot \left| \left\{ (i_1, \ldots, i_k) : i_1 \log q_1 + \cdots + i_k \log q_k < I \right\} \right|.$$ 

If the equations are linearly dependent, we could do away with fewer variables, but the linear dependence of the equations is difficult to analyse.

The easiest way to estimate the number of equations is via the following probabilistic argument. Let $X_1, \ldots, X_k$ be a sequence of independent random variables with $X_j$ taking the values $0, \ldots, n_j$ with equal probability. Fix a number $I \in \mathbb{R}_{\geq 0}$ Then the number of equations that the coefficients must satisfy so that $I_P(\alpha, \ldots, \alpha) \geq I$ divided by the number of coefficients is

$$P(X_1 \log q_1 + \cdots + X_k \log q_k \leq I).$$

So we can find a non-zero $P$ if the above quantity is less than $1/d$. If we choose $n_1, \ldots, n_k$ such that each of the terms are roughly of the
same order of magnitude, that is \( n_j \log q_j \) is roughly the same for each \( j \), then we can achieve

\[
P(X_1 \log q_1 + \ldots + X_k \log q_k \leq (2+\varepsilon)^{-1}(n_1 \log q_1 + \ldots + n_k \log q_k)) \leq d^{-1}
\]

by the law of large numbers, provided \( k \) is sufficiently large in terms of \( \varepsilon \). This means that we can achieve

\[
I_P(\alpha, \ldots, \alpha) > (2 + \varepsilon)^{-1}(n_1 \log q_1 + \ldots + n_k \log q_k),
\]

yielding

\[
(19) > q_1^{n_1} \cdots q_k^{n_k} = (17)
\]

completing our outline proof.

To prove Roth’s theorem, we need to carry out the following tasks.

- Construct a polynomial \( P \) with
  \[
  I_P(\alpha, \ldots, \alpha) \geq (2 + \varepsilon/10)^{-1}(n_1 \log q_1 + \ldots + n_k \log q_k).
  \]

  We will also need to control the size of the coefficients, so a simple argument based on the solubility of the system of linear equations will not suffice. If we solve the equations with, say, Cramer’s rule, then the solutions will be too large. However, it turns out, that it is possible to find much smaller solutions if the number of variables significantly exceeds (say, twice as many is more than enough) the number of equations. This goes by the name of Siegel’s Lemma, even though the argument, which is based on the box principle, has already been used by Thue.

- Show that \( P(p_1/q_1, \ldots, p_k/q_k) \) does not vanish. Unfortunately, there is no reason to expect that this holds. And indeed, if we add just one more equation to our system, we can, in fact arrange for \( P(p_1/q_1, \ldots, p_k/q_k) = 0 \). However, what is possible to show instead is that \( P \) cannot have large index at \( (p_1/q_1, \ldots, p_k/q_k) \). This is as good as if \( P(p_1/q_1, \ldots, p_k/q_k) \) did not vanish, because we can exchange \( P \) for a suitable derivative. By doing this, we decrease the index at \( (\alpha, \ldots, \alpha) \), but by not too much, so we can afford it. This goes by the name of Roth’s Lemma and it should be thought of as a converse for Siegel’s Lemma. What we show is that if \( q_1, \ldots, q_k \) are large and \( P \) have large index at \( (p_1/q_1, \ldots, p_k/q_k) \) (plus some further conditions are satisfied) than \( P \) must have large coefficients, larger than the bound we obtain in Siegel’s Lemma above. This is very easy to do for polynomials of a single variable, but very much harder if we have several variables. As we said above, this is the heart of the proof.

- After the above, all that remains is working out a rigorous upper bound using Taylor’s formula and the estimates for the coefficients of \( P \).
Before we move on, we point out some important technical aspects of the notation $P_{i_1, \ldots, i_k}$ and the fact that we divide by the factorials $i_1! \cdots i_k!$ in its definition. One convenient feature of this is that we do not have to write the factorials in Taylor’s formula. However, there is more to it. If $P$ has integer coefficients, then $P_{i_1, \ldots, i_k}$ also has integer coefficients. In fact, the coefficient of $x_1^{m_1-i_1} \cdots x_k^{m_k-i_k}$ in $P_{i_1, \ldots, i_k}$ is

\[
\binom{m_1}{i_1} \cdots \binom{m_k}{i_k}
\]
times the coefficient of $x_1^{n_1} \cdots x_k^{n_k}$ in $P$. This leads to

\[
H(P_{i_1, \ldots, i_k}) \leq 2^{n_1+\cdots+n_k} H(P),
\]
where $n_j$ is the degree of $P$ in $x_j$. If we did not divide by the factorials, we could have a weaker bound with $n_1! \cdots n_k!$ in place of $2^{n_1+\cdots+n_k}$. This is significant, because $2^{n_1+\cdots+n_k}$ will always be much smaller than $q_j^{n_j}$ provided $q_j$ is sufficiently large in term of $k$ and all the $q_j^{n_j}$ are of roughly the same size, and this is uniform in $n_j$. However, we cannot say the same about $n_1! \cdots n_k!$.

3.1. Some historical remarks. Thue’s proof can be fit in the scheme that we discussed above, by using polynomials of the form $P(x_1, x_2) = R(x_1) - Q(x_1)x_2$. However, Thue did not put it this way; more on this a bit later. Siegel used general polynomials in two variables, and probably he was the first one to understand the above scheme. The only thing that he was missing to obtain Roth’s theorem was Roth’s lemma.

Thue’s proof was based on the simple estimate

\[
\left| \frac{p}{q} - \frac{r}{s} \right| \geq \frac{1}{qs},
\]
which hold for any pair of distinct rational numbers. His plan was to construct suitable rational approximations of $\alpha$ at each scale, which would then repel away all other rational numbers. There are several methods that can produce approximations to a solution of a polynomial equation by iterating a rational function on a starting approximation. The first and most well known one is the Newton-Raphson method, but there are more refined ones. Starting with a good rational approximation to $\alpha$, we can produce a sequence of better and better rational approximations. These converge very fast to $\alpha$, but unfortunately, the denominators also grow very fast, so this will not work.

Instead, Thue looked for rational functions $R_n(x)/Q_n(x)$ such that $R_n(x)/Q_n(x) - \alpha$ vanishes at $x = \alpha$ to high order with $\deg R_n, \deg Q_n \leq n$. He constructed these using the box principle. He assumed that $\alpha$ has a very good approximation $p_0/q_0$, and then used the numbers

\[
\frac{p_n}{q_n} = \frac{R_n(p_0/q_0)}{Q_n(p_0/q_0)}
\]
to repel away the rationals.

Thue’s theorem says that

\[ |2^{1/3} - r/s| > \frac{c}{s^{2.5+\varepsilon}} \]

for some ineffective \( c \). To show that \( r/s \) is not violating this inequality, we need to find \( p/q \) such that

\[ |2^{1/3} - p/q| \leq \frac{1}{s^{2.5+\varepsilon}} \]

and \( q \leq s^{1.5+\varepsilon}/2 \) indeed, that would imply

\[ |2^{1/3} - r/s| \geq \frac{1}{qs} - \frac{1}{s^{2.5+\varepsilon}} \geq \frac{1}{s^{2.5+\varepsilon}}. \]

This looks very easy to satisfy. Indeed, we can do much better using just Dirichlet’s theorem. However, there is an important issue. We need to make sure that \( p/q \neq r/s \), and there is no way we can do that using Dirichlet’s theorem. On the other hand, Thue was able to show that in case \( R_n(p_0/q_0)/Q_n(p_0/q_0) = r/s \), we have

\[ \frac{(d^m/dx^m)R_n(p_0/q_0)}{(d^m/dx^m)Q_n(p_0/q_0)} \neq r/s \]

for some \( m \) that is not so large. This plays the role of Roth’s Lemma.

For a more complete and very nicely written discussion, see [16, Chapter 12].

3.2. **Siegel’s Lemma.** In this section, we look for small integer solutions of systems of linear equations. Before we state the main result, we need to introduce quantity that measures the complexity of the coefficients of the linear equations, because the size of the solution we can find will depend on this.

Let \( K \) be a number field, and let \( L = a_1x_1 + \ldots + a_Nx_N \in K[x_1, \ldots, x_N] \) be a linear form. The height of \( L \) is defined as

\[ H(L)_{[K:Q]} = \prod_v |L|_v = \prod_v \max(|a_1|_v, \ldots, |a_N|_v). \]

By the product formula, \( H(L) \) is invariant under multiplication by a non-zero element of \( K \).

**Siegel’s Lemma 33.** Let \( K \) be a number field and let \( D = [K : Q] \). Let \( M, N \in \mathbb{Z}_{\geq 0} \) with \( MD < N \), and let \( H \in \mathbb{R}_{\geq 1} \). Let \( L_1, \ldots, L_M \in K[x_1, \ldots, x_N] \) be linear forms with \( H(L_j) \leq H \) for all \( j \).

Then there are \( x_1, \ldots, x_N \in \mathbb{Z} \) not all 0 with

\[ L_j(x_1, \ldots, x_N) = 0, \quad j = 1, \ldots, M \]

and

\[ |x_j| \leq (N\mathcal{H})^{\frac{DM}{N-DM}}. \]
Some remarks are in order.

- This result is known as Siegel’s Lemma even though it is already contained implicitly in Thue’s work.
- The assumption $\mathcal{H} \geq 1$ is important. If we had e.g. $\mathcal{H} < 1/N$, then the conclusion would claim the existence of a non-zero integer with absolute value less than 1.
- Note that $DM$ is the number of linear equations over $\mathbb{Q}$ that the solution must satisfy.
- In a typical application of this lemma, we take $N$ to be a constant (larger than 1) multiple of $DM$, and we get an upper bound for the solution that is polynomial in $N$ and $\mathcal{H}$. In particular if $N \geq 2DM$, then we get $|x_j| \leq NH$.
- There is a refinement of Siegel’s Lemma due to Bombieri and Vaaler. They use the geometry of numbers instead of the box principle. For our purposes, the above bound will suffice, and we refer to the literature for the refinement. See for example [7, Section 2.9], or the original paper of Bombieri and Vaaler.

Before we discuss the proof of Siegel’s lemma we note the following corollary that is of interest to us.

**Corollary 34.** Let $\alpha \in \overline{\mathbb{Q}}$. Let $k \geq 1$, let $n_1, \ldots, n_k \in \mathbb{Z}_{\geq 1}$, let $w_1, \ldots, w_k \in \mathbb{R}_{>0}$ and let $I \in \mathbb{R}_{>0}$ be such that

$$\left| \{ (i_1, \ldots, i_k) \in \mathbb{Z}_{\geq 0}^k : i_1w_1 + \ldots + i_kw_k \leq I \} \right| \leq \left(\frac{(n_1 + 1) \cdots (n_k + 1)}{2[\mathbb{Q}(\alpha) : \mathbb{Q}]}\right)^k.$$

Then there is $P \neq 0 \in \mathbb{Z}[x_1, \ldots, x_k]$ of degree at most $n_j$ in $x_j$ such that

$$H(P) \leq (4H(\alpha))^{n_1 + \cdots + n_k}$$

and

$$I_P(\alpha, \ldots, \alpha; w_1, \ldots, w_k) \geq I.$$

**Proof.** We search for $P$ in the form

$$P = \sum_{j_1=0}^{n_1} \cdots \sum_{j_k=0}^{n_k} a_{j_1, \ldots, j_k} x_1^{j_1} \cdots x_k^{j_k}.$$

For $(i_1, \ldots, i_k) \in \mathbb{Z}_{\geq 0}^k$, we consider the linear form

$$L_{i_1, \ldots, i_k}(a_{j_1, \ldots, j_k}) = \sum_{j_1+\cdots+j_k=i_1+\cdots+i_k} \left(\frac{j_1}{i_1}\right) \cdots \left(\frac{j_k}{i_k}\right) \alpha^{j_1+\cdots+j_k-i_1-\cdots-i_k} a_{j_1, \ldots, j_k}.$$

Here the range of $j_l$ is from 0 to $n_l$. We note that $\left(\frac{j_l}{i_l}\right) = 0$ if $i_l > j_l$. The point of this is, of course, that $P_{i_1, \ldots, i_k} = 0$ is equivalent to

$$L_{i_1, \ldots, i_k}(a_{j_1, \ldots, j_k}) = 0.$$
We can estimate \(|L_{i_1,\ldots,i_k}|_v\) similarly to the proof of Proposition 29. For finite places \(v\), we write
\[
|L_{i_1,\ldots,i_k}|_v \leq \max(1, |\alpha|_v)^{n_1+\ldots+n_k}.
\]
For the infinite places \(v\), we write
\[
|L_{i_1,\ldots,i_k}|_v \leq 2^{n_1+\ldots+n_k} \max(1, |\alpha|_v)^{n_1+\ldots+n_k}.
\]
This is based on \((\tfrac{j}{n}) \leq 2^{-m}\) Multiplying together these bounds, we get
\[
H(L_{i_1,\ldots,i_k}) \leq (2H(\alpha))^{n_1+\ldots+n_k}.
\]
We can apply Siegel’s Lemma with \(D = [\mathbb{Q}(\alpha) : \mathbb{Q}], H = (2H(\alpha))^{n_1+\ldots+n_k}\) and
\[
N = (1 + n_1) \cdots (1 + n_k) \leq 2^{n_1+\ldots+n_k}.
\]
By the assumption of the corollary, the number \(M\) of linear forms we need to consider satisfies \(2DM \leq N\), and hence we obtain a non-zero solution with
\[
|a_{j_1,\ldots,j_k}| \leq NH \leq (4H(\alpha))^{n_1+\ldots+n_k}.
\]
This completes the proof. \(\square\)

Now we turn to the proof of Siegel’s Lemma. We will give the details only in the special case \(K = \mathbb{Q}\). The general case have the same proof, but the technicalities might obscure the ideas. After the proof, we will comment on what needs to be done differently in the general case.

**Proof of Siegel’s Lemma for \(K = \mathbb{Q}\).** We may assume without loss of generality that \(L_j\) has integral coefficients for all \(j\) and they have no common prime factors. Therefore, we have that the coefficients of \(L_j\) are bounded by \(H\).

Let
\[
Y = \left\lfloor (NH)^{\frac{M}{N}} \right\rfloor,
\]
and consider the vectors
\[(L_j(y_1,\ldots,y_N))_{j=1,\ldots,M}\]
for \(y_i\) running through 0, \ldots, \(Y\).

For each \(j\), using \(H(L_j) \leq H\), we get
\[
\max_{y_1,\ldots,y_N} L_j(y_1,\ldots,y_N) - \min_{y_1,\ldots,y_N} L_j(y_1,\ldots,y_N) \leq NHY.
\]
There are \((Y + 1)^N\) choices for \(y_1,\ldots,y_N\), and \(L_j(y_1,\ldots,y_N)\) may take at most \(NHY + 1 < NH(Y + 1)\) different values for each \(j\). (Here we used \(H \geq 1\) and \(N > M \geq 1\).) Note that
\[
(Y + 1)^N \geq (NH)^{\frac{MN-M}{N}} (Y + 1)^M = (NH(Y + 1))^M.
\]
By the box principle, there are two distinct vectors
\[(y_1,\ldots,y_N) \neq (z_1,\ldots,z_N) \in \{0,\ldots,Y\}^N\]
such that
\[ L_j(y_1, \ldots, y_N) = L_j(z_1, \ldots, z_N) \]
for all \( j \).

We observe that \( x_i = y_i - z_i \) for \( i = 1, \ldots, N \) is a solution with
\[ |x_i| \leq Y \leq (N\mathcal{H})^{\frac{M}{M-1}}. \]

\[ \square \]

We end this section by pointing out the additional ideas needed for the proof of Siegel’s Lemma for general number fields. For the full details, we refer to [16, Proposition 14.12]. We write \( \Sigma_{K,R} \) for the embeddings of \( K \) into \( \mathbf{R} \) and \( \Sigma_{K,C} \) for a collection of embeddings into \( \mathbf{C} \) which contain one from each pair of complex conjugates. We consider the map
\[ \Phi : K \to \mathbf{R}^{\Sigma_{K,R}} \times \mathbf{C}^{\Sigma_{K,C}} \equiv \mathbf{R}^D \]
\[ \alpha \mapsto (v(\alpha))_{v \in \Sigma_{K,R} \cup \Sigma_{K,C}}. \]

Similarly to the proof in the \( K = \mathbf{Q} \) case, we consider the vectors
\[ L_j(y_1, \ldots, y_N), \quad j = 1, \ldots, M \]
where \( y_i \) runs through \( 0, \ldots, Y \) for each \( i = 1, \ldots, N \), where \( Y \) is a fixed parameter. We can also show that the \( v \)-coordinate of \( \Phi(L_j(y_1, \ldots, y_N)) \) falls in an interval of length \( N|L_j|_v Y \) for each \( v \in \Sigma_{K,R} \cup \Sigma_{K,C} \). This allows us to confine the image of \( L_j(y_1, \ldots, y_N) \), \( j = 1, \ldots, M \) under \( \Phi \) in a certain box in \( (\mathbf{R}^{\Sigma_{K,R}} \times \mathbf{C}^{\Sigma_{K,C}})^M \).

Now we would like to have an upper bound on the number of possible vectors whose \( \Phi \) image falls in that box. To this end, we will use the product formula
\[ \prod_{v \in \Sigma_{K,R}} |\alpha|_v \cdot \prod_{v \in \Sigma_{K,C}} |\alpha|^2_v = \prod_{v \in \Sigma_{K,f}} |\alpha|_v^{-1} \]
for
\[ \alpha = L_j(y_1, \ldots, y_N) - L_j(z_1, \ldots, z_N). \]

For any \( \alpha \) of this form, we can write
\[ |\alpha|_v \leq |L_j|_v \]
for all finite place \( v \), so we get
\[ \prod_{v \in \Sigma_{K,R}} |\alpha|_v \cdot \prod_{v \in \Sigma_{K,C}} |\alpha|^2_v \geq \prod_{v \in \Sigma_{K,f}} |L_j|_v^{-1}. \]

This allows us to place a suitable box around the \( \Phi \) images of each
\[ L_j(y_1, \ldots, y_N), \quad j = 1, \ldots, M, \]
which will be pairwise disjoint. In fact, we can choose these boxes in many ways, (20) only requires that the product of the side lengths of the box belonging to the \( j \)th \( \mathbb{R}^{\Sigma_{K/R}} \times \mathbb{C}^{\Sigma_{K/C}} \) does not exceed \( \prod_{v \in \Sigma_{K_j}} |L_j|_V^{-1} \).

Now we can estimate the number of possible \( \Phi \) images of vectors of the form

\[ L_j(y_1, \ldots, y_N), \quad j = 1, \ldots, M \]

by comparing the volume of the big box containing all these vectors against the volume of the small boxes around each vector. By doing this, and choosing \( Y \) appropriately, the box principle will apply and we can find a solution for the system of linear equations in the same way we did in the special case \( K = \mathbb{Q} \).

3.3. Roth’s Lemma. The purpose of this section is to show that a polynomial \( P \in \mathbb{Z}[x_1, \ldots, x_k] \) with not too large coefficients and satisfying some further hypothesis cannot have large index at a rational point \( (p_1/q_1, \ldots, p_k/q_k) \).

The precise statement is the following.

**Roth’s Lemma 35.** For all \( \varepsilon > 0 \) and \( k \in \mathbb{Z}_{\geq 1} \) there is constant \( C = C(\varepsilon, k) \) such that the following holds. Let \( n_1, \ldots, n_k \in \mathbb{Z}_{\geq 1} \). Let \( p_1/q_1, \ldots, p_k/q_k \in \mathbb{Q} \) such that \( \log q_{j+1} \geq C \log q_j \) for \( j = 1, \ldots, k - 1 \) and

\[ \exp(n_1 + \ldots + n_k) \leq q_j^{n_j/C} \]

for all \( j \). Let \( P \neq 0 \in \mathbb{Z}[x_1, \ldots, x_k] \) be a polynomial of degree at most \( n_j \) in \( x_j \) for each \( j \). Suppose

\[ H(P) \leq q_j^{n_j/C} \]

for all \( j \). Then

\[ I_P(p_1/q_1, \ldots, p_k/q_k; \log q_1, \ldots \log q_k) \leq \varepsilon \sum n_j \log q_j. \]

Some remarks are in order.

- When we write down “let \( p/q \in \mathbb{Q} \)” we mean that \( \gcd(p, q) = 1 \). We have been ambiguous on this point until now, because it was not important.

- In a typical application of this lemma, the parameters are chosen in such a way that \( q_j^{n_j} \) is roughly of the same size for each \( j \), and \( q_1 \) is very large. Then the condition on \( H(P) \) is satisfied always when \( H(P) \leq \tilde{C}^{m_1+\ldots+n_k} \), where \( \tilde{C} \) can be taken arbitrarily large, if \( q_1 \) is large enough.

- Siegel’s Lemma implies that a polynomial violating the conclusion exists if we replace the bound on \( H(P) \) by \( H(P) \leq \max_j q_j^{C_2 n_j} \) for a suitably large \( C_2 = C_2(k) \). Therefore Roth’s lemma should be thought of as a converse to Siegel’s lemma.
Roth’s Lemma provides a converse to Siegel’s lemma in the aspect of the height of $P$. There is an alternative approach going back to Dyson. It is possible to show that a polynomial of certain degrees in $x_j$ cannot have large index both at $(\alpha, \ldots, \alpha)$ and at $(p_1/q_1, \ldots, p_k/q_k)$. This statement does not require any assumption on $H(P)$. In fact it can be formulated entirely in the setting of polynomials with complex coefficients. This has been done in the $k=2$ case by Dyson, and later it was later extended by others. For an introduction to this subject and to yet another method based of Faltings’ product theorem, we refer to Nakamaye’s survey [19].

The example $P(x_1, x_2) = (x_1 - x_2)^n$ shows that we need to make an assumption that excludes $p_1/q_1 = p_1/q_2$. This is achieved by the important assumption $\log q_2 \geq C \log q_1$. This may feel like an overkill, but there are other counterexamples, which we need to exclude. Consider for example $P = (R(x_1)x_2 - Q(x_1))^n$ for some $R, Q \in \mathbb{Z}[x_1]$, which vanishes to order $n$ at $x_1 = p_1/q_1$, $x_2 = Q(p_1/q_1)/R(p_1/q_1)$ for any choice of $p_1/q_1 \in \mathbb{Q}$. If $R$ and $Q$ has low degree compared to $\varepsilon$, this does not satisfy the conclusion of the lemma, but then $\log q_2 \geq C \log q_1$ also fails if $C$ is sufficiently large in terms of the degrees of $R$ and $Q$. We note that $P$ also vanishes at $(\alpha, \alpha)$ if $R(\alpha)\alpha - Q(\alpha) = 0$. However, this vanishing is of lower order than what we get when we apply Siegel’s lemma. Therefore, it is possible to exclude counterexamples of the above type by making a sufficiently stringent assumption on $I_P(\alpha, \alpha)$. The proof of Roth’s Lemma relies very heavily on the rapid growth of the $q_j$, and it is not clear to me if this condition can be relaxed if we combine it with other conditions.

Roth’s Lemma is proved by induction on $k$. We first explain the strategy. We begin with a simple auxiliary result that will help us to make computations with the index of a polynomial.

**Lemma 36.** Let $F, F^{(1)}, F^{(2)} \in \mathbb{Z}[x_1, \ldots, x_k]$, let $i_1, \ldots, i_k \in \mathbb{Z}_{\geq 0}$, let $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$, and let $w_1, \ldots, w_k \in \mathbb{R}_{>0}$. Then the following holds

$$I_{F_{i_1} \cdots i_k}(\alpha_1, \ldots, \alpha_k) \geq I_F(\alpha_1, \ldots, \alpha_k) - i_1 w_1 - \ldots - i_k w_k,$$

$$I_{F^{(1)}+F^{(2)}}(\alpha_1, \ldots, \alpha_k) \geq \min_{j=1,2} I_{F^{(j)}}(\alpha_1, \ldots, \alpha_k),$$

$$I_{F^{(1)}F^{(2)}}(\alpha_1, \ldots, \alpha_k) = \sum_{j=1,2} I_{F^{(j)}}(\alpha_1, \ldots, \alpha_k).$$

In these formulas the index is always understood to be with respect to $w_1, \ldots, w_k$. 
The first item follows easily from the definition of the index. The other two items can be deduced either by applying the rules of differentiation of sums and products or by considering Taylor expansions at $(\alpha_1, \ldots, \alpha_k)$. We omit the details.

The last item of the lemma suggests that Roth’s Lemma can be reduced to itself with $k-1$ in place of $k$ if $P$ can be factorized in the form

\[(21) \quad P(x_1, \ldots, x_k) = F(x_1, \ldots, x_{k-1})G(x_k).\]

Indeed, we could estimate the indices of $F$ and $G$ using Roth’s Lemma for lower values of $k$ and then the third item of Lemma 36 would complete the proof.

In general, we cannot hope to have a factorization of the above form. However, we can always have an identity of the form

\[(22) \quad P = F^{(1)}G^{(1)} + \ldots + F^{(h)}G^{(h)}\]

for some $h \in \mathbb{Z}_{\geq 1}$, $F^{(j)} \in \mathbb{Z}[x_1, \ldots, x_{k-1}]$ and $G^{(j)} \in \mathbb{Z}[x_k]$. Indeed, we could just write $P$ as a polynomial in $x_k$ with coefficients in $\mathbb{Z}[x_1, \ldots, x_{k-1}]$, and we get an identity of the above form.

Now the idea is to use the identity (22) to replace $P$ by another polynomial that admits a factorization of the form (21). To explore how this can be done, we first consider the simplest case $h = 2$. If the index of

\[F^{(1)}G^{(1)} + F^{(2)}G^{(2)}\]

is large at a point then the index of

\[F^{(1)} \frac{\partial}{\partial x_k} G^{(1)} + F^{(2)} \frac{\partial}{\partial x_k} G^{(2)}\]

is also large there. Now we can form a linear combination of these two functions to eliminate $F^{(2)}$. For example, we can multiply the first function by $(\partial/\partial x_k)G^{(2)}$ and subtract from it $G^{(2)}$ times the first one. We obtain

\[F^{(1)}(G^{(1)} \frac{\partial}{\partial x_k} G^{(2)} - \frac{\partial}{\partial x_k} G^{(1)} \cdot G^{(2)}).\]

By Lemma 36, this has also large index, and it factorizes in the required way.

We can do something similar also for general $h$. This leads us to the determinant

\[(23) \quad \begin{vmatrix} P & G^{(2)} & \ldots & G^{(h)} \\ \frac{\partial}{\partial x_k} P & \frac{\partial}{\partial x_k} G^{(2)} & \ldots & \frac{\partial}{\partial x_k} G^{(h)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^{h-1}}{\partial x_k^{h-1}} P & \frac{\partial^{h-1}}{\partial x_k^{h-1}} G^{(2)} & \ldots & \frac{\partial^{h-1}}{\partial x_k^{h-1}} G^{(h)} \end{vmatrix}.\]
Using Lemma 36, this has index as large as $\frac{h-1}{hx_k}P$, and using (22) for the first column, we get that (23) has the factorization

$$F_1 \cdot \left| \begin{array}{cccc} G_1 & G_2 & \cdots & G_h \\ \frac{\partial}{\partial x_k} G_1 & \frac{\partial}{\partial x_k} G_2 & \cdots & \frac{\partial}{\partial x_k} G_h \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^{h-1}}{\partial x_k^h} G_1 & \frac{\partial^{h-1}}{\partial x_k^h} G_2 & \cdots & \frac{\partial^{h-1}}{\partial x_k^h} G_h \end{array} \right|. \ \right|$$

This works very well when $h$ is small, but the degree becomes too large and the index too small when $h$ is large. It is better to consider a determinant all of whose entries involve a (not too high order) derivative of $P$.

Let $\mathcal{D}_k$ denote the set of differential operators of the form

$$\partial_{i_1, \ldots, i_k} := \frac{\partial^{i_1 + \cdots + i_k}}{i_1! \cdots i_k! \partial x_1^{i_1} \cdots \partial x_k^{i_k}}.$$

We call $i_1 + \cdots + i_k$ the order of $\partial_{i_1, \ldots, i_k}$. Let $\Delta_1, \ldots, \Delta_h \in \mathcal{D}_{k-1}$ be a sequence of differential operators, such that $\Delta_j$ is of order at most $j - 1$, and let

$$D_j = \frac{\partial^{j-1}}{(j-1)! \partial x_k^{j-1}}.$$

In the proof of Roth’s Lemma, we will consider the determinant

$$\left| \begin{array}{cccc} \Delta_1 D_1 P & \Delta_1 D_2 P & \cdots & \Delta_1 D_h P \\ \Delta_2 D_1 P & \Delta_2 D_2 P & \cdots & \Delta_2 D_h P \\ \vdots & \vdots & \cdots & \vdots \\ \Delta_h D_1 P & \Delta_h D_2 P & \cdots & \Delta_h D_h P \end{array} \right|,$$

which has a factorization of the form

$$\left| \begin{array}{cccc} \Delta_1 F^{(1)} & \Delta_1 F^{(2)} & \cdots & \Delta_1 F^{(h)} \\ \Delta_2 F^{(1)} & \Delta_2 F^{(2)} & \cdots & \Delta_2 F^{(h)} \\ \vdots & \vdots & \cdots & \vdots \\ \Delta_h F^{(1)} & \Delta_h F^{(2)} & \cdots & \Delta_h F^{(h)} \end{array} \right| \cdot \left| \begin{array}{cccc} D_1 G^{(1)} & D_2 G^{(1)} & \cdots & D_h G^{(1)} \\ D_1 G^{(2)} & D_2 G^{(2)} & \cdots & D_h G^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ D_1 G^{(h)} & D_2 G^{(h)} & \cdots & D_h G^{(h)} \end{array} \right|. \ \right|$$

The determinant on the right is called a Wronskian, which plays an important role in the theory of linear ODE’s. It is well known that the Wronskian of a sequence of linearly independent analytic functions does not vanish. We will need a version of this fact for generalized Wronskians of the type appearing on the left hand side of the factorization.

**Lemma 37.** Let $F^{(1)}, \ldots, F^{(h)} \in \mathbb{Q}[x_1, \ldots, x_k]$ be $\mathbb{Q}$-linearly independent polynomials. Then there is a sequence of differential operators
\[ \Delta_1, \ldots, \Delta_h \in D_k \text{ such that } \Delta_j \text{ is of order at most } j - 1 \text{ and } \begin{vmatrix} \Delta_1 F^{(1)} & \Delta_1 F^{(2)} & \cdots & \Delta_1 F^{(h)} \\ \Delta_2 F^{(1)} & \Delta_2 F^{(2)} & \cdots & \Delta_2 F^{(h)} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_h F^{(1)} & \Delta_h F^{(2)} & \cdots & \Delta_h F^{(h)} \end{vmatrix} \neq 0. \]

Some remarks are in order

- If \( k = 1 \), then the only possibility is that \( \Delta_j \) is \( j - 1 \)-fold differentiation in \( x_1 \). Indeed, there is no other variable to differentiate, and the \( \Delta_j \) has to be distinct, otherwise the determinant will have two identical rows and vanish. Hence, the lemma generalizes the above stated fact about the non-vanishing of Wronskians.

- In general, it may not be sufficient to take derivatives in any single variable. For example, consider the case \( k = 2 \), \( h = 4 \) and the functions \( 1, x, y, xy \). It can be seen that in this case the only non-vanishing generalized Wronskian will correspond to the differential operators \( 1, \partial/\partial x_1, \partial/\partial x_2 \) and \( \partial^2/\partial x_1\partial x_2 \).

We postpone the proof of this lemma and begin with the proof of Roth’s Lemma.

**Proof of Roth’s Lemma.** The proof is by induction on \( k \), and we first consider the \( k = 1 \) case. If \( p_1/q_1 \) is a root of multiplicity \( m \) of \( P \), then \( I_P(p_1/q_1) = m \log q_1 \). In addition, \( (q_1x - p_1)^m \) divides \( P \) in \( \mathbb{Z}[x] \), so the leading coefficient of \( P \) is divisible by \( q_1^m \). This means that

\[
H(P) \geq q_1^m = \exp(I_P(p_1/q_1; \log q_1)).
\]

The upper bound on \( H(P) \) assumed in the lemma yields

\[
I_P(p_1/q_1; \log q_1) \leq C^{-1}n_1 \log q_1,
\]

which proves the lemma with \( C = \varepsilon^{-1} \).

Now we consider the case \( k > 1 \). We fix a number \( \varepsilon_{k-1} \), which will be determined in the course of the proof, and it will be sufficiently small in terms of \( \varepsilon \) and \( k \). We assume that the lemma holds for \( k - 1 \) with \( \varepsilon_{k-1} \) and some \( C_{k-1} \).

We assume to the contrary that \( P \) and \( p_1/q_1, \ldots, p_k/q_k \) are counterexamples for the lemma for some \( C \) that will be chosen in the course of the proof, and it will be sufficiently large depending on \( \varepsilon, k, \varepsilon_{k-1} \) and \( C_{k-1} \).

We write \( P \) in the form

\[
P = F^{(1)}G^{(1)} + \cdots + F^{(h)}G^{(h)}
\]

for some \( 1 \leq h \leq n_k \). We assume, as we may, that \( F^{(1)}, \ldots, F^{(h)} \) and \( G^{(1)}, \ldots, G^{(h)} \) are \( \mathbb{Q} \)-linearly independent. Indeed, if this was not the case for the \( F^{(j)} \), say, then we could express one of them as a linear combination of the others and collecting the terms involving the other
$F^{(j)}$, we could have an identity of the form (24) with one less term. This process clearly has to terminate at some point.

We consider the matrix

$$
\mathcal{P} = \begin{bmatrix}
\Delta_1 D_1 P & \Delta_1 D_2 P & \ldots & \Delta_1 D_h P \\
\Delta_2 D_1 P & \Delta_2 D_2 P & \ldots & \Delta_2 D_h P \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_h D_1 P & \Delta_h D_2 P & \ldots & \Delta_h D_h P
\end{bmatrix},
$$

which has a factorization of the form

$$
\mathcal{F} \cdot \mathcal{G} = \begin{bmatrix}
\Delta_1 F^{(1)} & \Delta_1 F^{(2)} & \ldots & \Delta_1 F^{(h)} \\
\Delta_2 F^{(1)} & \Delta_2 F^{(2)} & \ldots & \Delta_2 F^{(h)} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_h F^{(1)} & \Delta_h F^{(2)} & \ldots & \Delta_h F^{(h)}
\end{bmatrix} \begin{bmatrix}
D_1 G^{(1)} & D_2 G^{(1)} & \ldots & D_h G^{(1)} \\
D_1 G^{(2)} & D_2 G^{(2)} & \ldots & D_h G^{(2)} \\
\vdots & \vdots & \ddots & \vdots \\
D_1 G^{(h)} & D_2 G^{(h)} & \ldots & D_h G^{(h)}
\end{bmatrix},
$$

where we choose the differential operators $\Delta_j \in \mathcal{D}_{k-1}$ of order at most $j - 1$ in such a way that $\mathcal{F} \neq 0$. This is possible by Lemma 37. Recall that

$$
D_j = \frac{\partial^{j-1}}{(j-1)! \partial x_k^{j-1}},
$$

and $\mathcal{G} \neq 0$ follows from Lemma 37.

We observe that $\mathcal{P} \in \mathbb{Z}[x_1, \ldots, x_k]$, $\mathcal{F} \in \mathbb{Z}[x_1, \ldots, x_{k-1}]$ and $\mathcal{G} \in \mathbb{Z}[x_k]$. In what follows, we estimate the degrees and heights of $\mathcal{F}$ and $\mathcal{G}$ in terms of the degrees and height of $P$. Then we will apply the $k - 1$ and $k = 1$ case of the lemma for $\mathcal{F}$ and $\mathcal{G}$ to obtain a bound for their indices at $(p_1/q_1, \ldots, p_{k-1}/q_{k-1})$ and $p_k/q_k$. This in turn will yield a bound on the index of $\mathcal{P}$ at $(p_1/q_1, \ldots, p_k/q_k)$, which will contradict a lower bound that we will obtain in terms of the index of $P$ at the same point.

The degree of each entry of $\mathcal{P}$ is at most $n_j$ in the variable $x_j$. This means that the degree of $x_j$ is at most $hn_j$ in $\mathcal{P}$, and therefore also in $\mathcal{F}$ for $j < h$ and in $\mathcal{G}$ for $j = h$.

The (naive) height of each entry of $\mathcal{P}$, is at most $2^{n_1 + \cdots + n_k} H(P)$. To calculate the determinant, we need to multiply together $h$ of these entries in $h!$ different ways and sum up the results. In each multiplication of $h$ entries, the monomials of the resulting polynomial is obtained by multiplying together $h$ monomials, one from each entry, and there are at most $((n_1 + 1) \cdots (n_k + 1))^h$ ways of choosing these monomials. We have therefore

$$
H(\mathcal{P}) \leq h^h ((n_1 + 1) \cdots (n_k + 1))^h (2^{n_1 + \cdots + n_k} H(P))^h \leq 2^{3hn_1 + \cdots + 3hn_k} H(P)^h.
$$

Since $\mathcal{F}$ and $\mathcal{G}$ has disjoint set of variables, we have $H(\mathcal{P}) = H(\mathcal{F}) H(\mathcal{G})$, and hence

$$
H(\mathcal{F}), H(\mathcal{G}) \leq 2^{3hn_1 + \cdots + 3hn_k} \cdot H(P)^h.
$$

Since $P$ satisfies the hypotheses of the lemma, we have

$$
H(\mathcal{F}), H(\mathcal{G}) \leq q_j^{\log(8)hn_j/C} \cdot q_j^{hn_j/C}
$$
for each \( j \). Provided \( C \) is large enough with respect to \( C_{k-1} \), we see that \( \mathcal{F} \) and \( \mathcal{G} \) satisfy the hypotheses of the lemma with \( n_j \) replaced by \( h n_j \), \( C \) replaced by \( C_{k-1} \) and \( \varepsilon \) replaced by \( \varepsilon_{k-1} \). Using the induction hypothesis, we have

\[
I_F(p_1/q_1, \ldots, p_k/q_k; \log q_1, \ldots, \log q_k) \leq \varepsilon_{k-1} \sum_{j=1}^{k-1} h n_j \log q_j
\]

\[
I_G(p_1/q_1, \ldots, p_k/q_k; \log q_1, \ldots, \log q_k) \leq \varepsilon_{k-1} h n_k \log q_k.
\]

By Lemma 36, we conclude

\[
I_P(p_1/q_1, \ldots, p_k/q_k; \log q_1, \ldots, \log q_k) \leq \varepsilon_{k-1} \sum_{j=1}^{k} h n_j \log q_j.
\]

It remains to give a lower bound on the index of \( \mathcal{P} \) in terms of the index of \( P \), which will contradict the above upper bound. By Lemma 36, the index of each entry in the first \( l \) columns of \( \mathcal{P} \) is at least

\[
\varepsilon \sum_{j=1}^{k} n_j \log q_j - (h - 1) \log q_{k-1} - (l - 1) \log q_k.
\]

If \( C \) is sufficiently large in terms of \( \varepsilon \), then \( \log q_{k-1} < \varepsilon/10 \cdot \log q_k \) and if \( l \leq \varepsilon/10 \cdot h + 1 \), then the above index is at least

\[
(\varepsilon/2) \sum_{j=1}^{k} n_j \log q_j.
\]

There are at least \( \varepsilon/10 \cdot h \) rows of \( \mathcal{P} \), where this bound holds. Now we expand the determinant \( \mathcal{P} \) and use Lemma 36 again to conclude

\[
I_P(p_1/q_1, \ldots, p_k/q_k; \log q_1, \ldots, \log q_k) \geq \varepsilon^2/20 \cdot \sum_{j=1}^{k} h n_j \log q_j.
\]

This clearly contradicts our upper bound provided \( \varepsilon_{k-1} \) is sufficiently small in terms of \( \varepsilon \). This completes the proof. \( \square \)

**Proof of Lemma 37.** First we prove the lemma in the case \( k = 1 \), and then we will reduce the general case to this in the second half of the proof.

We take \( \Delta_j = \partial^{j-1}/(j-1)! \partial x_1^{j-1} \). (As we discussed after the statement of the lemma, we do not have any other options in the univariate case.) By the properties of determinants and differentiation, non-vanishing of the Wronskian does not change if we replace \( F^{(j)} \) by \( a F^{(j)} - b F^{(i)} \) for some \( i \neq j \) and \( a, b \in \mathbb{Q} \) with \( a \neq 0 \). Using operations of this kind we can ensure that

\[
F^{(j)} = x_1^{m_j} + \text{lower order terms}
\]
for some \( m_j \) for each \( j \) and the \( m_j \) are distinct. If we show that
\[
\begin{vmatrix}
\Delta_1 x_1^{m_1} & \Delta_1 x_1^{m_2} & \ldots & \Delta_1 x_1^{m_h} \\
\Delta_2 x_1^{m_1} & \Delta_2 x_1^{m_2} & \ldots & \Delta_2 x_1^{m_h} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_h x_1^{m_1} & \Delta_h x_1^{m_2} & \ldots & \Delta_h x_1^{m_h}
\end{vmatrix}
\]
is non-zero, then this is the leading term of the determinant in the statement of the lemma, which hence does not vanish.

Thus it remains to show that (25) \( \neq 0 \). To this end, we write
\[
(25) = \begin{vmatrix}
\binom{m_1}{0} & \binom{m_2}{0} & \ldots & \binom{m_h}{0} \\
\binom{m_1}{1} & \binom{m_2}{1} & \ldots & \binom{m_h}{1} \\
\vdots & \vdots & \ddots & \vdots \\
\binom{m_1}{h-1} & \binom{m_2}{h-1} & \ldots & \binom{m_h}{h-1}
\end{vmatrix}
\times x_1^M,
\]
where \( M \in \mathbb{Z} \) is a number whose value is immaterial. The binomial coefficient \( \binom{m_j}{j} \) can be considered a polynomial of degree \( j \) in the variable \( m \). These are linearly independent polynomials, so no non-trivial linear combination of them for \( j = 0, \ldots, h-1 \) can vanish at \( h \) distinct points. This shows that the rows in the above determinant involving binomial coefficients are linearly independent, hence the determinant does not vanish.

This proves the lemma in the \( k = 1 \) case, and now we turn to the general case. We consider a rapidly growing sequence \( l_1, \ldots, l_h \) of integers and the homomorphism \( \Phi : \mathbb{Z}[x_1, \ldots, x_k] \to \mathbb{Z}[t] \) induced by \( x_j \mapsto t^{l_j} \). If the \( l_j \) are sufficiently fast increasing, then \( \Phi(F^{(j)}) \) for \( j = 1, \ldots, k \) are linearly independent. We assume that this is the case. From the \( k = 1 \) case, we can conclude that
\[
(26) = \begin{vmatrix}
D_1 \Phi(F^{(1)}) & D_1 \Phi(F^{(2)}) & \ldots & D_1 \Phi(F^{(h)}) \\
D_2 \Phi(F^{(1)}) & D_2 \Phi(F^{(2)}) & \ldots & D_2 \Phi(F^{(h)}) \\
\vdots & \vdots & \ddots & \vdots \\
D_h \Phi(F^{(1)}) & D_h \Phi(F^{(2)}) & \ldots & D_h \Phi(F^{(h)})
\end{vmatrix}
\]
does not vanish, where \( D_i = \frac{\partial}{\partial t} F^{(j)} (t^{l_1}, \ldots, t^{l_h}) = \sum_{a=1}^k \lambda_a t_1^{a-1} \frac{\partial}{\partial t^{a}} F^{(j)} (t^{l_1}, \ldots, t^{l_h}) \).

By the chain rule of differentiation, we have
\[
D_2 \Phi(F^{(j)}) = \frac{\partial}{\partial t} F^{(j)} (t^{l_1}, \ldots, t^{l_h}) = \sum_{a=1}^k \lambda_a t_1^{a-1} \frac{\partial}{\partial x_a} F^{(j)} (t^{l_1}, \ldots, t^{l_h}).
\]
Similarly, we have
\[
D_i \Phi(F^{(j)}) = \sum_{a_1, \ldots, a_k} A(t; a_1, \ldots, a_k) F^{(j)}_{a_1, \ldots, a_k} (t^{l_1}, \ldots, t^{l_h}),
\]
where the summation runs over non-negative integers satisfying \( a_1 + \ldots + a_k \leq i - 1 \), and \( A(t; a_1, \ldots, a_k) \) is a polynomial in \( t \) depending on
\(a_1, \ldots, a_k\). We plug these identities into (26). This yields a decomposition on (26) as a linear combination of the \(\Phi\) images of generalized Wronskians with coefficients in \(\mathbb{Z}[t]\). Since (26) \(\neq 0\), one of these generalized Wronskians must not vanish, and this proves the lemma. \(\square\)

3.4. Completing the proof. We complete the proof of Roth’s theorem. Let \(\varepsilon\) be as in the theorem. We first choose the value of the parameter \(k\). This is done in such a way that

\[
P(X_1 + \ldots + X_k \leq (1/2 - \varepsilon/10)k) \leq \frac{1}{10d},
\]

where \(X_j\) are independent uniformly distributed random variables in \([0, 1]\), and \(d = [Q(\alpha), Q]\). This is possible by the law of large numbers.

We assume to the contrary that there are infinitely many \(p/q \in \mathbb{Q}\) with \(|\alpha - p/q| \leq q^{-2-\varepsilon}\). We let \(p_1/q_1, \ldots, p_k/q_k\) be some of these so that the denominators are sufficiently large and they grow sufficiently rapidly so that the conditions of Roth’s lemma are satisfied. We will apply Roth’s lemma with \(\varepsilon/10\) in place of \(\varepsilon\) and some sufficiently large \(n_j\) satisfying that

\[
I\left(\frac{kY_j \log q_j}{n_1 \log q_1 + \ldots + n_k \log q_k}\right) \geq (1/2 - \varepsilon/5)(n_1 \log q_1 + \ldots + n_k \log q_k).
\]

Now we set \(n_1, \ldots, n_k\) in such a way that with independent random variables \(Y_j\) distributed uniformly in \([0, 1]\) we have

\[
P(Y_1 \log q_1 + \ldots + Y_k \log q_k \leq (1/2 - \varepsilon/10)(n_1 \log q_1 + \ldots + n_k \log q_k)) \leq \frac{1}{2d}.
\]

This can be achieved with an appropriate choice of the \(n_j\) if

\[
\frac{kY_j \log q_j}{n_1 \log q_1 + \ldots + n_k \log q_k}
\]

approximates \(X_j\) in distribution sufficiently well for each \(j\).

By Corollary 34 of Siegel’s Lemma, there is a polynomial \(P \in \mathbb{Z}[x_1, \ldots, x_k]\) of degree at most \(n_j\) in \(x_j\) with

\[
H(P) \leq (4H(\alpha))^{n_1 + \ldots + n_k}
\]

such that

\[
I_P(\alpha, \ldots, \alpha; \log q_1, \ldots, \log q_k) \geq (1/2 - \varepsilon/10)(n_1 \log q_1 + \ldots + n_k \log q_k).
\]

Now we apply Roth’s Lemma, and conclude that

\[
I_P(p_1/q_1, \ldots, p_k/q_k; \log q_1, \ldots, \log q_k) \leq (\varepsilon/10)(n_1 \log q_1 + \ldots + n_k \log q_k).
\]

This allows us to replace \(P\) by a suitable partial derivative that does not vanish at \((p_1/q_1, \ldots, p_k/q_k)\). We do this without changing notation. This comes with the small expense of replacing our previous height and index bounds by

\[
H(P) \leq (8H(\alpha))^{n_1 + \ldots + n_k}
\]

\[
I_P(\alpha, \ldots, \alpha; \log q_1, \ldots, \log q_k) \geq (1/2 - \varepsilon/5)(n_1 \log q_1 + \ldots + n_k \log q_k)
\]
By the non-vanishing of \( P(p_1/q_1, \ldots, p_k/q_k) \), we have the lower bound
\[
|P(p_1/q_1, \ldots, p_k/q_k)| \geq q_1^{-n_1} \cdots q_k^{-n_k}.
\]

We consider the Taylor expansion
\[
P(p_1/q_1, \ldots, p_k/q_k) = \sum_{i_1, \ldots, i_k} P_{i_1, \ldots, i_k}(\alpha, \ldots, \alpha) \prod_j (\alpha - p_j/q_j)^{i_j}
\]
with the aim of finding an upper bound. We have
\[
H(P_{i_1, \ldots, i_k}) \leq (16H(\alpha))^{n_1 + \cdots + n_k}.
\]
Using \(|\alpha| \leq H(\alpha)^d\), we can write
\[
|P_{i_1, \ldots, i_k}(\alpha, \ldots, \alpha)| \leq (32H(\alpha)^{d+1})^{n_1 + \cdots + n_k},
\]
where we used \((n_1 + 1) \cdots (n_k + 1) \leq 2^{n_1 + \cdots + n_k}\) to bound the number of terms in the polynomial. Using the same bound for the number of terms in the Taylor expansion, we can write
\[
|P(p_1/q_1, \ldots, p_k/q_k)| \leq (64H(\alpha)^{d+1})^{n_1 + \cdots + n_k} \exp(-(2 + \varepsilon)I_P(\alpha, \ldots, \alpha)) \leq (64H(\alpha)^{d+1})^{n_1 + \cdots + n_k} (q_1^{-n_1} \cdots q_k^{-n_k})^{(2+\varepsilon)(1/2-\varepsilon/5)}
\]
Comparing this to the lower bound, we get a contradiction if
\[
(64H(\alpha)^{d+1})^{n_1 + \cdots + n_k} < (q_1^{n_1} \cdots q_k^{n_k})^{\varepsilon/10-\varepsilon^2/5}.
\]
This is of the same shape as what we have assumed when we chosen \(q_j\). In any case, it certainly holds if each \(q_j\) are sufficiently large in terms of \(\varepsilon\) and \(\alpha\), so we could have certainly assumed it even if we have not. So we reached the desired contradiction and the proof is complete.

4. Baker’s theorem – proof

The purpose of this section is to prove the following weaker version of Baker’s theorem stated in the introduction. Our exposition is based on Baker’s book [5].

**Theorem 38.** Let \(\alpha_1, \alpha_2, \alpha_3 \in \overline{\mathbb{Q}}_{\neq 0}\) and let \(\log \alpha_1, \log \alpha_2\) and \(\log \alpha_3\) be any choices of their logarithms. Suppose \(2\pi i, \log \alpha_1, \log \alpha_2\) and \(\log \alpha_3\) are linearly independent over \(\mathbb{Q}\). Then \(\log \alpha_1, \log \alpha_2\) and \(\log \alpha_3\) are linearly independent over \(\overline{\mathbb{Q}}\).

This statement is weaker than what we stated earlier in several respects. First, it is about three logarithms rather than arbitrarily many. We made this restriction only to simplify our notation in the proof; it does not affect the arguments. Second and third, we assume more than just the independence of \(\log \alpha_1, \log \alpha_2\) and \(\log \alpha_3\) (over \(\mathbb{Q}\)), and we conclude less than the independence of \(1, \log \alpha_1, \log \alpha_2\) and \(\log \alpha_3\) (over \(\overline{\mathbb{Q}}\)). These two changes simplify the argument in a significant way. After we finished the proof we will comment on how the proof needs to be changed to obtain the stronger form we stated earlier.
Now we discuss the proof strategy. We will prove the contrapositive. That is, we assume that $\log \alpha_1$, $\log \alpha_2$ and $\log \alpha_3$ are linearly dependent over $\mathbb{Q}$, so without loss of generality we can assume the existence of $\beta_1, \beta_1 \in \mathbb{Q}$ such that
\begin{equation}
\log \alpha_3 = \beta_1 \log \alpha_1 + \beta_2 \log \alpha_2,
\end{equation}
and we prove that $2\pi i, \log \alpha_1, \log \alpha_2$ and $\log \alpha_3$ are linearly dependent over $\mathbb{Q}$. To this end, it will be enough to show that the numbers
$$
\alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}
$$
for $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{Z}$ are not all distinct.

In the proof, we will work with functions of the form
\begin{equation}
F(z) = \sum_{\lambda_1,\lambda_2,\lambda_3=0}^{\Lambda-1} p(\lambda_1, \lambda_2, \lambda_3) \alpha_1^{\lambda_1 z} \alpha_2^{\lambda_2 z} \alpha_3^{\lambda_3 z},
\end{equation}
where $\Lambda \in \mathbb{Z}_{\geq 1}$ is a parameter, which we will choose later, and $p(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{Z}$ for all $\lambda_1, \lambda_2, \lambda_3$. The notation requires some explanation. When we write down $\alpha_1^z$ for some $z \in \mathbb{C}$, we mean $\exp(\log \alpha_1 \cdot z)$, where $\log \alpha$ is the choice of the logarithm we made in the statement of the theorem.

If we are able to find such a function that vanishes at the points $z = 1, \ldots, L^3$, then writing $\gamma_1, \ldots, \gamma_{\Lambda^3}$ for an enumeration of the numbers $\alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}$ we see that the determinant
$$
\begin{vmatrix}
\gamma_1 & \ldots & \gamma_{\Lambda^3} \\
\vdots & \ddots & \vdots \\
\gamma_{\Lambda^3} & \ldots & \gamma_{\Lambda^3}
\end{vmatrix}
$$
vanishes. Indeed, the linear combination of the columns with the coefficients $p(\lambda_1, \lambda_2, \lambda_3)$ is the column vector with entries $F(1), \ldots, F(L^3)$, which vanishes. The above determinant can only vanish if at least two of the $\gamma_j$ are equal, which is precisely what we would like to show.

The numbers $\alpha_1^{\lambda_1 z} \alpha_2^{\lambda_2 z} \alpha_3^{\lambda_3 z}$ for $z \in \mathbb{Z}$ are algebraic, so we can use Siegel’s lemma to construct a function of the form (28) that vanishes at many integer points. The number of zeros will be less than the required $\Lambda^3$, but we hope to increase the number of zeros of $F$ using the following idea. Since we know that $F$ vanishes at $z = 1, \ldots, L_0$ for some $L_0$, an argument based on complex analysis will yield a very good upper bound for $F$ at $z = L_0 + 1, L_0 + 2, \ldots$. We will combine this with height bounds on $F(z)$, which is algebraic for $z \in \mathbb{Z}$, and conclude that $F$ must vanish on a longer range of the integers. If necessary, we repeat this argument until we show that $F$ must vanish for $z = 1, \ldots, \Lambda^3$.

It turns out that this argument requires higher order vanishing of $F$ at $z = 1, \ldots, L_0$. This causes some difficulty, because $F'(z)$ and higher derivatives do not take algebraic values at integer points. Indeed, differentiating $\exp(\log \alpha_j \cdot z)$ we introduce $\log \alpha_j$ factors. The solution to this issue is that we can write $F'(z)$ as a linear combination collecting
all terms that involve the same powers of \( \log \alpha_j \) and use Siegel’s Lemma to find the coefficients \( p(\lambda_1, \lambda_2, \lambda_3) \) such that each term in this linear combination vanishes.

The neatest way to keep track of the \( \log \alpha_j \) factors is by introducing a function in 3 complex variables. We write

\[
\Phi(z_1, z_2, z_3) = \sum_{\lambda_1, \lambda_2, \lambda_3=1}^{\Lambda-1} p(\lambda_1, \lambda_2, \lambda_3) \alpha_1^{\lambda_1} z_1 \alpha_2^{\lambda_2} z_2 \alpha_3^{\lambda_3}.
\]

Using the notation

\[
\Phi_{m_1, m_2, m_3} = \frac{\partial^{m_1+m_2+m_3} \Phi}{\partial z_1^{m_1} \partial z_2^{m_2} \partial z_3^{m_3}},
\]

(which differs from the notation we used in the proof of Roth’s theorem), we can write

\[
\frac{d^m}{dz^m} F(z) = \sum_{m_1, m_2, m_3 \geq 0 \atop m_1 + m_2 + m_3 = m} \Phi_{m_1, m_2, m_3}(z, z, z)
\]

and for \( z \in \mathbb{Z} \),

\[
(\log \alpha_1)^{-m_1} (\log \alpha_2)^{-m_2} (\log \alpha_3)^{-m_3} \Phi_{m_1, m_2, m_3}(z, z, z)
\]

is a linear combination of \( p(\lambda_1, \lambda_2, \lambda_3) \) with algebraic coefficients.

Now we make an attempt to implement our general plan. We proceed with a series of informal statement, which we will make precise later after some necessary modification, which will become clear later.

**Almost Lemma.** There is a choice of not too large integers \( p(\lambda_1, \lambda_2, \lambda_3) \) for \( \lambda_1, \lambda_2, \lambda_3 = 1, \ldots, \Lambda \), not all zero, such that

\[
\Phi_{m_1, m_2, m_3}(l, l, l) = 0
\]

for all \( l = 1, \ldots, L_0 \) and \( m_1, m_2, m_3 \geq 0 \) with \( m_1 + m_2 + m_3 < M_0 \) provided

\[
\Lambda^3 \geq CM_0^3 L_0,
\]

where \( C \) is a constant that depends on \( \alpha_1, \alpha_2, \alpha_3 \).

The only thing that is imprecise in this statement is the bound on \( p(\lambda_1, \lambda_2, \lambda_3) \), which we ignore for now. This statement is proved using Siegel’s lemma. There are \( \Lambda^3 \) variables, which need to satisfy less than \( M_0^3 \) constraints over \( \mathbb{Q}(\alpha_1, \alpha_2, \alpha_3) \). If we choose \( C \) sufficiently large, the number of variables exceeds twice the number of constraints over \( \mathbb{Q} \), so Siegel’s lemma applies and it gives a good bound on the size of the solution in terms of the heights of the constraints.

Next we give simple upper and lower bounds for \( \Phi_{m_1, m_2, m_3} \). The upper bound will not make use of the zeros. That will be exploited later on to improve the initial upper bound, which we give now.
Almost Lemma. There is constant $C = C(\alpha_1, \alpha_2, \alpha_3)$ such that the function $\Phi$ constructed in the previous statement satisfies

$$|\Phi_{m_1, m_2, m_3}(z, z, z)| \leq C|z|^{A_1 + M_0 \log \Lambda}$$

$$|\Phi_{m_1, m_2, m_3}(l, l, l)| \geq C^{-l \Lambda - M_0 \log \Lambda}$$

for all $m_1, m_2, m_3 \geq 0$ with $m_1 + m_2 + m_3 < M_0$, $z \in \mathbb{C}$ and $l \in \mathbb{Z}_{\geq 1}$.

This statement is again imprecise, because it ignores the contribution of the coefficients $p(\lambda_1, \lambda_2, \lambda_3)$, so it is a bit stronger than what can be proved. The upper bound follows simply from the triangle inequality. Indeed, the terms in the definition of $\Phi$ are of size $\max(1, |\alpha_1 \alpha_2 \alpha_3|)|z|^2$ if we ignore $p(\lambda_1, \lambda_2, \lambda_3)$, and the $C M_0 \log \Lambda$ factor comes in when we differentiate $\Phi$. The lower bound can be deduced from height bounds of similar nature and the general bound $|\beta| \geq H(\beta) - \deg \beta$.

The next statement is precise, and we will prove it later. It is a simple application of the maximum modulus principle. It can be used together with the upper bound in the previous statement together with information about the zeros of $\Phi_{m_1, m_2, m_3}(z, z, z)$ to produce an even better upper bound.

Lemma 39. Let $A_1, A_2 \in \mathbb{R}_{>0}$, and let $L, M \in \mathbb{Z}_{>0}$ such that $M > 2A_2$. Let $f$ be an entire function, that is a function that is holomorphic on $\mathbb{C}$. Suppose $|f(z)| \leq A_1 \cdot \exp(A_2|z|)$, and $f$ has zeros of order at least $M$ at $z = 1, \ldots, L$. Then

$$|f(u)| \leq A_1 \left(\frac{2eA_2}{LM}\right)^L$$

for all $L \leq u \leq LM/A_2$.

Now we take stock of what we have done so far. We apply Lemma 39 for the function $\Phi_{m_1, m_2, m_3}(z, z, z)$ for some $m_1, m_2, m_3$ satisfying say $m_1 + m_2 + m_3 \leq M_0/2$. By the almost lemmata, we know that the hypotheses of the lemma are satisfied with $L = L_0$, $M = M_0/2$, $A_2 = C\Lambda$ and some $A_1$. Now we would like to take $u = L_0 + 1$ and obtain an upper bound that is smaller that the lower bound in the second almost lemma with $l = u$. To this end, we would need roughly speaking, say

$$\frac{CL_0 \Lambda}{L_0 M_0} = \frac{2eA_2}{LM} < 1/2$$

and

$$L_0 M_0/2 = LM > L_0 \Lambda,$$

and we also need to satisfy the condition $M_0 > 2A_2 = C\Lambda$ of the lemma. All three of these boil down to the requirement that $\Lambda$ is much smaller than $M_0$. Unfortunately, this cannot be satisfied, because we need to satisfy the requirement $\Lambda^3 \geq C M_0^3 L_0$ in the first almost lemma.
This failure, as disappointing as it might be, is certainly not unexpected. We have never made use of the assumption (27), without which we certainly cannot conclude the proof. Using this identity, we can express the logarithmic factors in terms of \( \log \alpha_1 \) and \( \log \alpha_2 \) alone. If we do this and separate the terms involving \((\log \alpha_1)^m (\log \alpha_2)^n\) in the application of Siegel’s lemma, then we will have \( CM_0^2 L_0 \) constraints rather than \( CM_0^3 L_0 \). This will allow us to choose a smaller \( \Lambda \), which is precisely what we need for the argument to succeed.

In the next section, we implement this change and carry out the full details of the argument.

4.1. The details. In this section, we have the following standing notation. We let \( \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2 \in \mathbb{Q} \) with \( \alpha_j \neq 0 \), and we let \( \log \alpha_j \) be any fixed choice of the logarithm of \( \alpha_j \). We assume the identity

\[
\log \alpha_3 = \beta_1 \log \alpha_1 + \beta_2 \log \alpha_2.
\]

We fix a number \( \Lambda \in \mathbb{Z} \geq 1 \) and consider the function

\[
\Phi(z_1, z_2) = \sum_{\lambda_1, \lambda_2, \lambda_3 = 0}^{\Lambda-1} p(\lambda_1, \lambda_2, \lambda_3) \alpha_1^{\lambda_1+\beta_1 \lambda_3} z_1^1 \alpha_2^{\lambda_2+\beta_2 \lambda_3} z_2^2
\]

for some \( p(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{Z} \). We note that

\[
\Phi_{m_1, m_2}(z, z) = (\log \alpha_1)^{m_1} (\log \alpha_2)^{m_2} \times \sum_{\lambda_1, \lambda_2, \lambda_3 = 0}^{\Lambda-1} p(\lambda_1, \lambda_2, \lambda_3) (\lambda_1 + \beta_1 \lambda_3)^{m_1} (\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1 z} \alpha_2^{\lambda_2 z} \alpha_3^{\lambda_3 z}
\]

for all \( m_1, m_2 \in \mathbb{Z}_{\geq 0} \). For integer values of \( z \),

\[
(\log \alpha_1)^{-m_1} (\log \alpha_2)^{-m_2} \Phi_{m_1, m_2}(z, z)
\]

is a linear combination of \( p(\lambda_1, \lambda_2, \lambda_3) \) with algebraic coefficients.

Now we give suitably modified and precise versions of the statements we gave above and prove them.

**Lemma 40.** With our standing notation, there is a constant \( C \) depending only on \( \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2 \) such that the following holds. Let \( M_0, L_0, \Lambda \in \mathbb{Z}_{>0} \) be such that \( \Lambda^3 \geq CM_0^2 L_0 \).

Then there are \( p(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{Z} \) for \( \lambda_1, \lambda_2, \lambda_3 = 0, \ldots, \Lambda - 1 \), not all zero, such that

\[
P := \max_{\lambda_1, \lambda_2, \lambda_3} |p(\lambda_1, \lambda_2, \lambda_3)| \leq C L_0^{\Lambda} A + M_0 \log \Lambda
\]

and

\[
(29) \quad \Phi_{m_1, m_2}(l, l) = 0
\]

for all \( l = 1, \ldots, L_0 \) and \( m_1, m_2 \geq 0 \) with \( m_1 + m_2 < M_0 \).
Proof. The condition (29) is equivalent to $p(\lambda_1, \lambda_2, \lambda_3)$ being a solution of the system of linear equations

$$\sum_{\lambda_1, \lambda_2, \lambda_3} p(\lambda_1, \lambda_2, \lambda_3) (\lambda_1 + \beta_1 \lambda_3)^{m_1} (\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3} = 0$$

for all $l = 1, \ldots, L_0$ and $m_1, m_2 \geq 0$ with $m_1 + m_2 < M_0$.

We use Siegel’s Lemma to find a small solution of this system. We have $\Lambda_0$ variables and $(M_0)L$ equations. If $C$ is sufficiently large in terms of $D = [Q(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2) : Q]$, then $\Lambda_0 \geq 2D(M_0)L$. We assume that this is the case.

It remains to estimate the height of the equations. For a finite place $v$, we can write

$$|(\lambda_1 + \beta_1 \lambda_3)^{m_1} (\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}|_v \leq \max(1, |\beta_1|_v)^{M_0} \max(1, |\beta_2|_v)^{M_0} \max(1, |\alpha_1|_v)^{\Lambda_0} \max(1, |\alpha_2|_v)^{\Lambda_0}.$$ 

For infinite places, we can write

$$|(\lambda_1 + \beta_1 \lambda_3)^{m_1} (\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}|_v \leq (CA)^{M_0} C^{\Lambda_0}.$$ 

Multiplying these together, we find that the heights of the equations are bounded by $C^{\Lambda_0 + M_0 \log A}$.

Now Siegel’s Lemma can be applied, and it yields a solution with

$$\max |p(\lambda_1, \lambda_2, \lambda_3)| \leq \Lambda_0 \cdot C^{\Lambda_0 + M_0 \log A}.$$ 

This completes the proof, because the $\Lambda_0$ factor can be absorbed into the second factor by making $C$ larger.

\[\square\]

**Lemma 41.** With our standing notation and with $p(\lambda_1, \lambda_2, \lambda_3)$ as in Lemma 40, there is a constant $C$ depending only on $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$, such that

$$H((\log \alpha_1)^{-m_1} (\log \alpha_2)^{-m_2} \Phi_{m_1, m_2}(l, l)) \leq C^{(L_0 + l)\Lambda + M_0 \log A}$$

for all $l \in \mathbb{Z}_{>0}$ and $m_1, m_2 \geq 0$ with $m_1 + m_2 < M_0$. In particular, if $\Phi_{m_1, m_2}(l, l) \neq 0$, we have

$$|\Phi_{m_1, m_2}(l, l)| \geq C^{-(L_0 + l)\Lambda - M_0 \log A}$$

for the same ranges of $l, m_1, m_2$.

In addition, we have

$$|\Phi_{m_1, m_2}(z, z)| \leq C^{(L_0 + |z|)\Lambda + M_0 \log A}$$

for the same ranges of $m_1, m_2$ and all $z \in \mathbb{C}$.

**Proof.** We recall that

$$(\log \alpha_1)^{-m_1} (\log \alpha_2)^{-m_2} \Phi_{m_1, m_2}(l, l) = \sum_{\lambda_1, \lambda_2, \lambda_3} p(\lambda_1, \lambda_2, \lambda_3) (\lambda_1 + \beta_1 \lambda_3)^{m_1} (\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}.$$
As above, for finite places, we can write

\[
\left| \sum_{\lambda_1, \lambda_2, \lambda_3} p(\lambda_1, \lambda_2, \lambda_3)(\lambda_1 + \beta_1 \lambda_3)^{m_1}(\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1 l} \alpha_2^{\lambda_2 l} \alpha_3^{\lambda_3 l} \right|_v \\
\leq \max(1, |\beta_1|_v)^{M_0} \max(1, |\beta_2|_v)^{M_0} \max(1, |\alpha_1|_v)^{\Lambda L_0} \\
\times \max(1, |\alpha_2|_v)^{\Lambda L_0} \max(1, |\alpha_3|_v)^{\Lambda L_0}.
\]

For the infinite places, we write

\[
\left| \sum_{\lambda_1, \lambda_2, \lambda_3} p(\lambda_1, \lambda_2, \lambda_3)(\lambda_1 + \beta_1 \lambda_3)^{m_1}(\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1 l} \alpha_2^{\lambda_2 l} \alpha_3^{\lambda_3 l} \right|_v \\
\leq \Lambda^3 \cdot P \cdot (CL)^{2M_0} C^{|\lambda|}.
\]

Multiplying together these estimates and plugging in the bound for \(P\) from Lemma 40, we get the claim about height.

Combining the height bound with Lemma 30, we get the claimed lower bound on \(|\Phi_{m_1, m_2}(l, l)|\).

Finally, the upper bound on \(|\Phi_{m_1, m_2}(z, z)|\) follows from (30), which is also valid when we replace \(l\) by an arbitrary complex number.

\[\square\]

**Lemma 42.** Let \(A_1, A_2 \in \mathbb{R}_{>0}\), and let \(L, M \in \mathbb{Z}_{>0}\) such that \(M > 2A_2\). Let \(f\) be an entire function, that is a function that is holomorphic on \(\mathbb{C}\). Suppose \(|f(z)| \leq A_1 \cdot \exp(A_2|z|)\), and \(f\) has zeros of order at least \(M\) at \(z = 1, \ldots, L\). Then

\[|f(u)| \leq A_1 \left(\frac{2euA_2}{LM}\right)^L\]

for all \(L \leq u \leq LM/A_2\).

**Proof.** The function

\[\frac{f(z)}{(z-1)^M \cdots (z-L)^M}\]

is entire. Fix a number \(R \geq 2L\). By the maximum modulus principle, we have

\[|f(z)| \leq \frac{A_1 \cdot \exp(A_2R)}{(R/2)^{LM}}\]

for all \(z\) with \(|z| \leq R\). Here we used \(|z - l| \geq R/2\) for \(l = 1, \ldots, L\).

For all \(u\) with \(L \leq u \leq R\) and \(l = 1, \ldots, L\), we have \(|u - l| \leq u\).

Therefore

\[|f(u)| \leq A_1 \exp(A_2R) \cdot \left(\frac{2u}{R}\right)^{LM}.
\]

Now we set \(R = LM/A_2\). This satisfies our requirement \(R \geq 2L\) by the assumption \(M > 2A_2\). We conclude

\[|f(u)| \leq A_1 e^{LM} \cdot \left(\frac{2uA_2}{LM}\right)^L,
\]

as required.

\[\square\]

Now we combine the above lemmata in the following statement.
Proposition 43. With our standing notation and with \( p(\lambda_1, \lambda_2, \lambda_3) \) as in Lemma 40, there is a constant \( C = C(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2) \) such that the following holds. Let \( L \geq L_0 \) and \( M \leq M_0 \) be integers such that
\[
2|M, \\
M > C\Lambda, \\
LM > CM_0 \log \Lambda.
\]
Suppose \( \Phi_{m_1, m_2}(l, l) = 0 \) for \( m_1, m_2 \geq 0 \) with \( m_1 + m_2 \leq M \) and \( l = 1, \ldots, L \).

Then \( \Phi_{m_1, m_2}(l, l) = 0 \) for \( m_1, m_2 \geq 0 \) with \( m_1 + m_2 \leq M/2 \) and \( l = 1, \ldots, \lfloor LM/C\Lambda \rfloor \).

Proof. By the upper bound in Lemma 41, we can apply Lemma 42 for \( f(z) = \Phi_{m_1, m_2}(z, z) \) for any \( m_1, m_2 \geq 0 \) with \( m_1 + m_2 \leq M/2 \) with
\[
A_1 = C^{L_0 \Lambda + M_0 \log \Lambda}, \quad A_2 = C\Lambda
\]
and \( M/2 \) in place of \( M \). We note that the condition \( M > 2A_2 \) holds by the assumptions of the proposition, and we also have
\[
\frac{2elA_2}{LM/2} < \frac{1}{2},
\]
for \( l \in \mathbb{Z} \cap [L, LM/C\Lambda] \), provided \( C \) is sufficiently large. For these \( l \), we get
\[
|f(l)| \leq C^{L_0 \Lambda + M_0 \log \Lambda} 2^{M/2}.
\]

Again, by the assumptions of the proposition, we have
\[
C^{L_0 \Lambda + M_0 \log \Lambda} 2^{M/2} < C^{-(L_0 + l)\Lambda - M_0 \log \Lambda},
\]
and so we can conclude
\[
f(l) = \Phi_{m_1, m_2}(l, l) = 0,
\]
as required. \( \square \)

Now we complete the proof of Baker’s theorem. We choose \( L_0 \) to be a sufficiently large power of 2 and take (say) \( M_0 = L_0^2 \). This allows us to set \( \Lambda \leq C(M_0^2 L_0)^{1/3} \leq CL_0^{5/3} \) for a suitably large \( C \) (independent of \( L_0 \)) so that Lemma 40 applies.

This means in particular that
\[
\Phi_{m_1, m_2}(l, l) = 0
\]
for all \( m_1, m_2 \geq 0 \) with \( m_1 + m_2 < M_0 \) and \( l = 1, \ldots, L_0 \).

We apply Proposition 43 repeatedly so that after the \( k \)’th application we have
\[
\Phi_{m_1, m_2}(l, l) = 0
\]
for all $m_1, m_2 \geq 0$ with $m_1 + m_2 < M_0/2^k$ and $l = 1, \ldots, L_k$, where $L_k$ is recursively defined by

$$L_k = \left[ L_{k-1} M/2^{k-1} C \Lambda \right].$$

We note that $M$ is a power of 2 by definition, so the first condition of Proposition 43 is satisfied as long as $k \leq \log M_0 / \log 2$. The second condition is satisfied provided

$$2^k < C M_0 / \Lambda \leq C L_0^{1/3}.$$

We observe that $L_k M/2^k$ is monotone increasing, so the last condition of the proposition is always satisfied if

$$L_0 M_0 > C M_0 \log \Lambda,$$

which holds provided $L_0$ is sufficiently large.

This means that we can apply the proposition at least $C^{-1} \log L_0$ times. We note we have

$$L_k \geq c(k) L_0 (M_0 / \Lambda)^k \geq C^{-1} c(k) L_0^{1+k/3}$$

for some $c(k) > 0$ depending only on $k$. If $L_0$ is sufficiently large, this gives $L_{13} > \Lambda^3$, and this proves the theorem, as we explained at the beginning of the section.

4.2. **Refinements of the argument.** In this section, we informally discuss a number of refinements of the argument, which yield various improvements of the result.

4.2.1. *Weakening the hypothesis.* It is enough to assume in the theorem that $\log \alpha_1$, $\log \alpha_2$ and $\log \alpha_3$ are linearly independent over $\mathbb{Q}$, there is no need to include $2\pi i$.

To achieve this improvement in our proof of the contrapositive, we need to find a vanishing linear combination of $\log \alpha_1$, $\log \alpha_2$ and $\log \alpha_3$ not involving $2\pi i$ at the end of the argument.

Baker’s papers contain two different ways of achieving this. The first one is based on the observation that the function

$$0 \neq F(z) = \sum_{\lambda_1, \lambda_2, \lambda_3} p(\lambda_1, \lambda_2, \lambda_3) \alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}$$

may have a zero of order $\Lambda^3$ at $z = 0$ only if the numbers $\lambda_1 \log \alpha_1 + \lambda_2 \log \alpha_2 + \lambda_3 \log \alpha_3$ are not all distinct. Indeed, the vector

$$(F(0), \frac{d}{dz} F(0), \ldots, \frac{d^{A^3-1}}{dz^{A^3-1}} F(0))$$

is a linear combination with coefficients $p(\lambda_1, \lambda_2, \lambda_3)$ of vectors of the form

$$(\lambda_1 \log \alpha_1 + \lambda_2 \log \alpha_2 + \lambda_3 \log \alpha_3)^0, \ldots, (\lambda_1 \log \alpha_1 + \lambda_2 \log \alpha_2 + \lambda_3 \log \alpha_3)^{A^3-1},$$

which are the rows of a Vandermonde determinant.
So it would be enough to show that $F$ has a zero of order $\Lambda^3$ at $z = 0$. Our argument yields that $F$ vanishes at $z = 1, \ldots, \Lambda^{100}$, say and even to a very high order. This in turn implies that $F$ is very small in a fairly large disk around 0 using an argument similar to the proof of Lemma 42. Now we can combine this with Cauchy’s formula for $(d^m/dz^m)F(0)$ applied for a suitable circle around 0 to show that $(d^m/dz^m)F(0)$ must be very small. Unfortunately, we are not able to conclude that $(d^m/dz^m)F(0) = 0$, for this number is not algebraic.

However, we can still conclude that the vectors (31) have a linear combination with integral coefficients, which is very small. Now it follows that the Vandermonde determinant must be very small, too, and we conclude that

$$ (\lambda_1 - \tilde{\lambda}_1) \log \alpha_1 + (\lambda_2 - \tilde{\lambda}_2) \log \alpha_2 + (\lambda_3 - \tilde{\lambda}_3) \log \alpha_3 $$

must be very small for some $(\lambda_1, \lambda_2, \lambda_3) \neq (\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3)$. If (32) vanishes, we have exactly what we want. If (32) does not vanish, then

$$ \alpha_1^{\lambda_1 - \tilde{\lambda}_1} \alpha_2^{\lambda_2 - \tilde{\lambda}_2} \alpha_3^{\lambda_3 - \tilde{\lambda}_3} $$

is an algebraic number very close but not equal to 1. By computing the height of this number minus 1, we can see that this is not possible.

The second argument to achieve the same outcome involves showing that

$$ F(1/K) = F(2/K) = \ldots = F(\Lambda^3/K) = 0 $$

for our auxiliary function $F$ defined above and for a suitably large $K$ depending on $\log \alpha_1, \log \alpha_2, \log \alpha_3$ and $\Lambda$. Using our arguments, we can show that $F$ is very small at these points. Now these numbers are also algebraic and their height will not be too large if $K$ is only moderately large. Therefore, we can conclude that $F$ indeed vanishes at these points.

Now we can conclude that

$$ \frac{\lambda_1}{K} \log \alpha_1 K = \frac{\lambda_2}{K} \log \alpha_2 K = \frac{\lambda_3}{K} \log \alpha_3 K $$

for some $(\lambda_1, \lambda_2, \lambda_3) \neq (\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3)$ in the same way as before. This implies that

$$ \frac{\lambda_1 - \tilde{\lambda}_1}{K} \log \alpha_1 + \frac{\lambda_2 - \tilde{\lambda}_2}{K} \log \alpha_2 + \frac{\lambda_3 - \tilde{\lambda}_3}{K} \log \alpha_3 $$

is an integral multiple of $2\pi i$. If $K$ is sufficiently large, then this integer multiple must be 0.

4.2.2. Inhomogeneous forms. It is possible to strengthen the conclusion of Theorem 38 by claiming the linear independence of 1, $\log \alpha_1$, $\log \alpha_2$ and $\log \alpha_3$ over $\mathbb{Q}$ and not just the linear independence of the latter three numbers.
In our proof of the contrapositive, this means that we start with an inhomogeneous linear identity of the form
\[
\beta_0 + \beta_1 \log \alpha_1 + \beta_2 \log \alpha_2 = \log \alpha_3
\]
with some \(\beta_0, \beta_1, \beta_2 \in \overline{\mathbb{Q}}\).

This requires us to modify our auxiliary function and set
\[
\Phi(z_1, z_2, z_3) = \sum_{\lambda_0, \ldots, \lambda_3=0}^{\Lambda-1} p(\lambda_0, \ldots, \lambda_3)z_0^{\lambda_0} e^{\lambda_0 \lambda_3 z_0} \alpha_1^{(\lambda_1 + \beta_1 \lambda_3)z_1} \alpha_2^{(\lambda_2 + \beta_2 \lambda_3)z_2}.
\]

The key properties of this function are that
\[
\Phi(z, z, z) = \sum_{\lambda_0, \ldots, \lambda_3=0}^{\Lambda-1} p(\lambda_0, \ldots, \lambda_3)z^{\lambda_0} \alpha_1^{\lambda_1 z} \alpha_2^{\lambda_2 z} \alpha_3^{\lambda_3 z},
\]
the partial derivatives
\[
(\log \alpha_1)^{-m_1}(\log \alpha_1)^{-m_2}\Phi_{m_0, m_1, m_2}(l, l, l)
\]
only involve algebraic numbers for integral values of \(l\), and the number of parameters \(\lambda_0, \ldots, \lambda_3\) exceeds the number of variables \(z_0, \ldots, z_2\).

These properties are everything needed for the argument, and we can conclude in the same way that \(\Phi(l, l, l)\) vanishes at \(l = 1, \ldots, \Lambda^4\).

To conclude the proof, we need to show that the vectors
\[
(l^{\lambda_0}(\alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}))_{l=1, \ldots, \Lambda^4}
\]
are linearly independent provided the numbers \(\alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3}\) are all distinct. (This is enough only to prove the theorem with the stronger hypothesis, but using the ideas discussed in the previous section, we can obtain the theorem with the weaker hypothesis.)

The desired linear independence follows from the following result.

**Lemma 44.** Let \(\gamma_1, \ldots, \gamma_N\) be distinct non-zero complex numbers and let \(D \in \mathbb{Z}_{\geq 0}\). The \(DN \times DN\) determinant whose columns are indexed by the pair \(i = 0, \ldots, D - 1, j = 1, \ldots, N\) and are given by
\[
[\begin{bmatrix}l^{\lambda_0}(\alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3})\end{bmatrix}_{l=1, \ldots, DN}]^T
\]
does not vanish.

**Sketch of proof.** We consider the determinant as a polynomial in \(\gamma_1\). By selecting the highest degree terms from the columns involving \(\gamma_1\), it is possible to show that the degree of this polynomial is at most
\[
DN + \ldots + (DN - D + 1).
\]
It is also possible to show that this polynomial vanishes at \(\gamma_1 = 0\) to order at least
\[
1 + \ldots + D,
\]
and at each \(\gamma_1 = \gamma_j\) for \(j > 1\) to order at least \(D\).
A moment of reflection gives

\[ DN + \ldots + (DN - D + 1) = 1 + \ldots + D + D(N - 1), \]

so we conclude that the determinant in the statement of the lemma equals

\[ A \gamma_1^{1+\ldots+D} \prod_{j>2} (\gamma_1 - \gamma_j)^D, \]

where \( A \) is the leading coefficient of the above discussed polynomial.

Considering the terms in the expansion of the determinant involving the highest possible powers of \( \gamma_1 \), we see that the leading coefficient is the product of a Vandermonde determinant (which arises as a minor of the columns involving \( \gamma_1 \)) and a determinant of the same form as that in the lemma (which arises as a minor of the columns not involving \( \gamma_1 \)) but with \( N - 1 \) in place of \( N \). By induction, we can conclude the proof. □

For a full proof, see [5, Lemma 3.3]. The above statement says that a function of the form

\[ \sum_{j=1}^N P_j(z) \gamma_j^l, \]

where \( P_j \) is a polynomial of degree at most \( D - 1 \), cannot vanish simultaneously at \( l = 1, \ldots, ND \). The number of zeros of such functions in various regions of the complex plane has been studied extensively. Upper bounds for them are known as “zero estimates” and play an important role (just like Roth’s lemma and its extensions) in the theory.

### 4.3. Lower bounds for linear forms in logarithms

The above argument can be used with simple modifications to show lower bounds for linear forms in logarithms of the kind

\[ \beta_1 \log \alpha_1 + \beta_2 \log \alpha_2 + \beta_3 \log \alpha_3. \]

For simplicity, we may assume that \( \beta_3 = -1 \). If (33) is small, then we have the approximate identity

\[ \log \alpha_3 \approx \beta_1 \log \alpha_1 + \beta_2 \log \alpha_2. \]

Using Siegel’s Lemma, just as we did in Lemma 40, we can find \( p(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{Z} \) such that

\[ \sum_{\lambda_1, \lambda_2, \lambda_3} p(\lambda_1, \lambda_2, \lambda_3)(\lambda_1 + \beta_1 \lambda_3)^{m_1}(\lambda_2 + \beta_2 \lambda_3)^{m_2} \alpha_1^{\lambda_1} \alpha_2^{\lambda_2} \alpha_3^{\lambda_3} \]

vanishes in an appropriate range of \( l \) and \( m_1, m_2 \).

Just as we did in Lemma 41, we can give lower bounds on sums of the form (34), when it does not vanish (so in particular the parameters are outside the range of the first lemma). Moreover, we can also give an upper bound for the partial derivatives \( \Phi_{m_1, m_2}(z, z) \) of the function

\[ \Phi(z_1, z_2) = \sum_{\lambda_1, \lambda_2, \lambda_3=0}^{\lambda-1} p(\lambda_1, \lambda_2, \lambda_3) \alpha_1^{(\lambda_1+\beta_1 \lambda_3)z_1} \alpha_2^{(\lambda_2+\beta_2 \lambda_3)z_2} \]
similar to those in Lemma 41.

The crucial difference is that when we try to apply Lemma 42 for the functions \( \Phi_{m_1, m_2}(z, z) \), they will not vanish at \( l = 1, \ldots, L_0 \), but instead, we only know that it has small derivatives at these points up to a certain order. So we need a version of Lemma 42 with a relaxed hypothesis. Informally speaking, let \( f \) be an entire function such that \((d^m/dz^m)f(l)\) is small for \( l = 1, \ldots, L \) and \( m = 0, \ldots, M - 1 \). We would like to obtain an upper bound for \( f(u) \) for some \( u \geq L \). To this end, we can write

\[
\frac{f(u)}{(u-1)^M \cdots (u-L)^M} = \frac{1}{2\pi i} \int_{|z|=R} \frac{f(z)}{(z-1)^M \cdots (z-L)^M (z-u)} dz + \text{Residue terms.}
\]

We can estimate the integral by estimating the integrand pointwise, and we get a similar bound to what we had when we applied the maximum modulus principle in Lemma 42. The residue terms can be estimated in terms of \((d^m/dz^m)f(l)\).

The rest of the argument can be carried out as we did above, and it yields that if (33) is small enough, then \( \log \alpha_1, \log \alpha_2 \) and \( \log \alpha_3 \) are linearly dependent over \( \mathbb{Q} \). Such a linear relation can be used to eliminate one of the logarithms and then we can run the same argument again with two logarithms instead of three. At the end of the day, this argument gives a lower bound for non-vanishing expressions of the form (33).

This argument has been refined by Baker and other authors (the contribution of Feldman is especially important) to obtain better and better bounds. For accounts of the history of these developments we refer to the literature, for which good starting points are the books [4], [9] and [23].

5. Complexity of digital expansion of algebraic numbers

In this coming part of the course, we discuss further applications of the subspace theorem and lower bounds on linear forms in logarithms. We begin with an application of the former.

Let \( \alpha \in [0, 1] \) be a number and let \( 0.\alpha_1\alpha_2 \ldots \) be its base \( b \) expansion for some integer \( b \in \mathbb{Z}_{\geq 2} \). We define the complexity function of the digit expansion of \( \alpha \) as

\[
\rho_\alpha(n) = |\{\alpha_k \alpha_{k+1} \ldots \alpha_{k+n-1} : k \geq 1\}|,
\]

that is the number of different substrings of the digit expansion of length \( n \).

It is easy to prove that \( \rho_\alpha(n) = b^n \) for almost all \( \alpha \) for all \( b \) and \( n \), and it is very reasonable to expect that this also holds for irrational algebraic numbers. However, this problem appears to be a long way
beyond what existing methods can achieve. The best known result towards it is due to Adamczewski and Bugeaud [1].

**Theorem 45.** All irrational algebraic numbers \( \alpha \in [0, 1] \) satisfy

\[
\lim_{n \to \infty} \frac{\rho_n(n)}{n} = \infty.
\]

This result implies in particular that numbers generated by an automaton are either rational or algebraic. For an explanation of what this means, we refer to the paper of Adamczewski and Bugeaud [1] or the survey of Bilu [6], where an exposition of Theorem 45 is given along with a string of other striking applications of the subspace theorem.

The results of Adamczewski and Bugeaud are more general, in particular, they have similar results when the base of the expansion is a Pisot or Salem number. (Again for definitions and details, we refer to the original papers.)

The proof of the theorem relies on a transcendence criterion, which is based in turn on the subspace theorem.

To state this transcendence criterion, we need the following definition. We say that (the digit expansion of) a number \( \alpha \) has long repetitions if there is \( \varepsilon > 0 \) such that there are infinitely many \( n \in \mathbb{Z}_{>0} \) such that \( \alpha_1 \ldots \alpha_n \) contains two identical substrings of length at least \( \varepsilon n \), i.e. if there are \( k, m, l \) with \( 1 \leq k < m \leq n - l + 1 \) and \( l \geq \varepsilon n \) such that

\[
\alpha_k = \alpha_m, \ldots, \alpha_{k+l-1} = \alpha_{m+l-1}.
\]

**Theorem 46** (Adamczewski, Bugeaud, Luca). Let \( \alpha \in [0, 1] \) and \( b \in \mathbb{Z}_{\geq 2} \). If the digit expansion of \( \alpha \) contains long repetitions, then \( \alpha \) is either rational or transcendental.

It clear from the pigeon hole principle that Theorem 46 implies Theorem 45.

For integer bases \( b \), the Theorem 46 is due to Adamczewski, Bugeaud and Luca [2]. It has been generalized to Pisot and Salem bases by Adamczewski and Bugeaud [1]. We stated the result in a slightly stronger form than in the literature, because our definition of long repetitions does not require the two occurrences of the repeated substrings to be disjoint. A very simple combinatorial argument shows that the repeated substrings can always be taken disjoint at the expense of making the repetition slightly shorter. However, disjointness is not needed for the proof of Theorem 46, so it is not necessary to discuss this combinatorial argument.

We recall the subspace theorem in a slightly stronger form than how we stated it before.

**Theorem 47** (*S*-adic subspace theorem, Schlickewei). Let \( d \in \mathbb{Z}_{\geq 2} \). Let \( S \) be a finite set of places of \( \mathbb{Q} \) containing \( \infty \). Let \( K \) be a number field, and for each \( v \in S \), fix an extension of \( | \cdot |_v \) from \( \mathbb{Q} \) to \( K \).
to $K$. For each $v \in S$, let $L_1^{(v)}, \ldots, L_d^{(v)} \in K[x_1, \ldots, x_d]$ be linearly independent linear forms. Then for all $\varepsilon > 0$, the solutions of

$$\prod_{v \in S} \prod_{j=1}^d |L_j^{(v)}(x_1, \ldots, x_d)|_v \leq H(x_1, \ldots, x_d)^{-\varepsilon}$$

lie in a finite union of proper subspaces of $\mathbb{Q}^d$.

Proof of Theorem 46. Suppose that $\alpha$ is algebraic and its digit expansion contains long repetitions. We show that $\alpha$ is rational.

Let $\varepsilon > 0$ be suitably small for the definition of long repetitions applied to $\alpha$, and let $k, m, n, l$ be such that $1 \leq k \leq m \leq n - l + 1$, $l \geq \varepsilon n$ and

$$\alpha_k = \alpha_m, \ldots, \alpha_{k+l-1} = \alpha_{m+l-1}.$$}

Observe that the digit expansion of the fractional parts of $b^k \alpha$ and $b^m \alpha$ begin with identical strings of length $l$. This means that these two numbers are very close to each other in $\mathbb{R}/\mathbb{Z}$. More precisely, there is an integer $A$ such that

$$|b^m \alpha - b^k \alpha - A| \leq b^{-l} \leq b^{-\varepsilon n}.$$}

Now we set up the linear forms for which we will apply the subspace theorem. We let $S$ consist of the prime divisors of $b$ and $\infty$, and let $d = 3$. For a finite $v \in S$, we set

$$L_j^{(v)} = X_j, \quad j = 1, 2, 3,$$

and we set

$$L_1^{(\infty)} = X_1, \quad L_2^{(\infty)} = X_2, \quad L_3^{(\infty)} = \alpha X_1 - \alpha X_2 - X_3.$$}

We observe that

$$\prod_{v \in S} \prod_{j=1}^3 |L_j^{(v)}(b^m, b^k, A)|_v \leq b^{-\varepsilon n},$$

and

$$H(b^m, b^k, A) \leq b^n.$$}

By the subspace theorem, there is a finite union of proper subspaces of $\mathbb{Q}^3$, which is independent of $k, l, m, n$, and which contains $(b^m, b^k, A)$. We claim that only the subspace $L_3^{(\infty)} = 0$ may contain infinitely many such triplets. For every given $\alpha$, there are infinitely many choices for $(b^m, b^k, A)$, so one of them must satisfy

$$\alpha b^m - \alpha b^k - A = 0,$$

and this implies that $\alpha$ is rational.

Now we prove the claim. Let $V$ be a proper subspace of $\mathbb{Q}^3$ that contains infinitely many triples $(b^m, b^k, A)$ satisfying (35) and (36). We must have $\dim V \geq 2$, unless one of the linear forms vanishes on $V$. (Recall what we said about lines containing infinitely many solutions.)
The only linear form that may vanish on infinitely many solutions is \( L_3^{(\infty)} \). So we may assume \( \dim V = 2 \).

We apply the subspace theorem again on \( V \). For each place \( v \), we select two among \( L_1^{(v)} \), \( L_2^{(v)} \) and \( L_3^{(v)} \) whose restriction to \( V \) are linearly independent. This is possible, because the restrictions of these linear forms certainly span the dual of \( V \). Moreover, we may include \( L_3^{(\infty)} \), unless \( V = \ker(L_3^{(\infty)}) \). Indeed, if \( L_3^{(\infty)}|_V \neq 0 \), then it cannot be proportional to both \( L_1^{(\infty)}|_V \) and \( L_2^{(\infty)}|_V \).

Therefore, we can apply the subspace theorem again if \( V \neq \ker(L_3^{(\infty)}) \). In that case, we find that a proper subspace of \( V \) contains infinitely many triples \((b^m, b^k, A)\), which is not possible, as we have seen above. Thus, we must have \( V = \ker(L_3^{(\infty)}) \), as required.

6. THE ORDER \( \times 2, \times 3 \) IN \( \mathbb{Z}/q\mathbb{Z} \)

The purpose of this section is to prove the following result, which is another application of the subspace theorem.

**Theorem 48.** For an integer \( q \in \mathbb{Z}_{\geq 1} \) with \( \gcd(q, 6) = 1 \), write \( o(q) \) for the order of the multiplicative subgroup of \( \mathbb{Z}/q\mathbb{Z} \) generated by 2 and 3. Then

\[
\lim_{q \to \infty} \frac{o(q)}{(\log q)^2} = \infty.
\]

This follows very easily from the following result, which was proved independently by Corvaja and Zannier [12], and Hernández and Luca [13] after initial work by Bugeaud, Corvaja and Zannier [10].

**Theorem 49** (Corvaja, Zannier; Hernández, Luca). Write \( S \) for the set of numbers of the form \( 2^n3^m \) for \( n, m \in \mathbb{Z}_{\geq 0} \). Then for all \( \varepsilon > 0 \), there are only finitely many pairs of multiplicatively independent \( a, b \in S \) such that

\[
\gcd(a - 1, b - 1) \geq \max(a, b)^\varepsilon.
\]

It was observed by Bugeaud, Corvaja and Zannier that there are infinitely many \( n \in \mathbb{Z} \) such that

\[
\gcd(2^n - 1, 3^n - 1) \geq 3^{n^{\varepsilon/\log \log n}}.
\]

Indeed, there are infinitely many integers \( n \) such that the number of distinct primes \( p \) with \( p - 1 | n \) is at least

\[
p^{\varepsilon/\log \log n}.
\]

Each such prime \( p \) divides \( \gcd(2^n - 1, 3^n - 1) \), and their product is larger than

\[
3^{n^{\varepsilon/\log \log n}}.
\]

The claim about the existence of such integers \( n \) is proved in [3, Proposition 10]. We only include a vague heuristic to suggest why it is reasonable to expect that it holds. Fix some number \( x \), and let \( n \) be
the product of all primes between \( x/2 \) and \( x \). There are approximately \( cx/\log x \) such primes, so \( n \) is approximately \( x^{cx/\log x} = e^{cx} \). This means that \( n \) has \( c \log n/\log \log n \) prime factors, and hence it has \( 2^{c \log n/\log \log n} \) divisors. It is reasonable to expect that at least a \( c/\log n \) proportion of these divisors \( d \) will be such that \( d + 1 \) is a prime.

We begin by deducing the first theorem from the second.

**Proof of Theorem 48.** Let

\[
\Lambda = \{(n, m) \in \mathbb{Z}^2 : 2^n 3^m \equiv 1 \mod q \}.
\]

Then \( \Lambda \) is a lattice inside \( \mathbb{Z}^2 \) and \( [\mathbb{Z}^2 : \Lambda] = o(q) \).

Let \((n, m) \neq (0, 0) \in \Lambda \). Then either \( q|2^n - 3^m| \neq 0 \) or \( q|2^n - 3^m - 1 \neq 0 \) depending on the signs of \( n \) and \( m \). This means that we must have

\[
2^n 3^m > q,
\]

hence \( \max(|n|, |m|) \geq \log_3 q/2 \).

By standard results in the geometry of numbers, e.g. Minkowski’s second theorem, there are linearly independent \((n_1, m_1), (n_2, m_2) \in \Lambda \) with

\[
\max(|n_1|, |m_1|) \max(|n_2|, |m_2|) \leq C o(q)
\]

for some absolute constant \( C \). In light of the above lower bound, we have

\[
\max(|n_j|, |m_j|) \leq C \frac{o(q)}{\log q}
\]

for \( j = 1, 2 \).

In addition, the angle of \((n_1, m_1)\) and \((n_2, m_2)\) is bounded below by a universal constant and their lengths are within a factor of \( C o(q)/(\log q)^2 \). This means that we can replace \((n_1, m_1)\) and \((n_2, m_2)\) by their linear combinations with integer coefficients so that we have \( n_j, m_j > 0 \), \((n_1, m_1)\) and \((n_2, m_2)\) are still linearly independent, and we still have

\[
\max(|n_j|, |m_j|) \leq C \frac{o(q)}{\log q}
\]

for \( j = 1, 2 \).

Now we set \( a = 2^n 3^m \) and \( b = 2^n 3^m \), and observe

\[
\gcd(a - 1, b - 1) \geq q.
\]

We note that

\[
\max(a, b) \leq q^{o(q)/(\log q)^2},
\]

hence

\[
\gcd(a - 1, b - 1) \geq \max(a, b) \frac{c(\log q)^2}{o(q)}.
\]

By Theorem 49, we must have

\[
\lim_{q \to \infty} \frac{(\log q)^2}{o(q)} = 0,
\]

which proves our claim. \(\square\)
Now we turn to the proof of Theorem 49, which is based on the subspace theorem. It requires the following well known result, which we will prove later using the subspace theorem again.

**Proposition 50.** Let \( L \in \mathbb{Q}[x_1, \ldots, x_d] \) be a linear form. Then there is a constant \( C = C(L) \) such that
\[
L(x_1, \ldots, x_d) = 0, \quad x_1, \ldots, x_d \in S
\]
implies that
\[
|x_i - x_j|_\infty |x_i - x_j|_2 |x_i - x_j|_3 \leq C
\]
for some \( i \neq j \).

The construction of the linear forms in the following proof is due to Levin [15].

**Proof of Theorem 49.** Fix \( \varepsilon > 0 \). Let \( a, b \in S \) be multiplicatively independent and let \( d = \gcd(a - 1, b - 1) \). Suppose that \( d \geq \max(a, b)^\varepsilon \).

We will show that \( d \) must be bounded above by a constant depending on \( \varepsilon \).

We fix a suitable large \( n \in \mathbb{Z}_{\geq 1} \). We will apply the subspace theorem on the vector space
\[
V := \mathbb{Q}^n/\{(y, \ldots, y) : y \in \mathbb{Q}\}.
\]

We will evaluate our linear forms on the point
\[
(e_1/d, \ldots, e_{n^2}/d),
\]
where \( e_1, \ldots, e_{n^2} \) is an enumeration of the numbers \( a^i b^j \) for \( i = 0, \ldots, n - 1 \) and \( j = 0, \ldots, n - 1 \) such that \( e_1 = 1 \) and \( e_{n^2} = a^{n-1} b^{n-1} \). Note that \( e_j \equiv 1 \mod d \) for all \( j \), hence \( e_i/d - e_j/d \in \mathbb{Z} \) for all \( i, j \).

We set \( S = \{\infty, 2, 3\} \). To apply the subspace theorem, we select a basis \( L_1^{(v)}, \ldots, L_{n^2-1}^{(v)} \) of \( V^* \) for each \( v \) in such a way that
\[
\prod_{j=1}^{n^2-1} |L_j^{(v)}(e_1/d, \ldots, e_{n^2}/d)|_v
\]
is small.

In addition, we will also fix a basis \( x_1, \ldots, x_{n^2-1} \) of \( V^* \) such that
\[
x_i(e_1/d, \ldots, e_{n^2}/d) \in \mathbb{Z}
\]
for each \( i \). Then we can express each \( L_j^{(v)} \) as a linear combination of the \( x_i \), and these will be the linear forms for which we apply the subspace theorem. The choice of the \( x_i \) does not affect the proof except that it has to be chosen independently of \( a \) and \( b \). So for concreteness we can set, say, \( x_i = y_i - y_{n^2} \), where \( y_1, \ldots, y_{n^2} \) are the coordinates on \( \mathbb{Q}^{n^2} \).
We choose each $L_j^{(v)}$ in the form $y_k - y_l$. A simple analysis shows that there is no advantage of considering more general forms. We note the inequalities

$$|e_k/d - e_l/d|_v \leq \max(|e_k/d|_v, |e_l/d|_v).$$

If $v$ is finite, this is just the ultra metric triangle inequality. If $v = \infty$, then it follows because both $e_k$ and $e_l$ are non-negative.

Now we set the linear forms $L_j^{(\infty)}$. We note that $|e_n/d|_\infty$ is minimal and $|e_1/d|_\infty$ is maximal among the $|e_j/d|_\infty$. For simplicity of discussion, suppose that the $e_j/d$ are in increasing order of $|\cdot|_\infty$. There must be one linear form among the $L_j^{(\infty)}$ that involves $y_{n^2}$. Also, there must be one among the rest that involves at least one among $y_{n^2-1}$ and $y_{n^2}$, and similar observations hold for the rest of the forms. This shows that the best upper bound we can hope for on

$$n^2 - 1 \prod_{j=1}^{n^2-2} |L_j^{(\infty)}(e_1/d, \ldots, e_{n^2}/d)|_\infty$$

is

$$\prod_{j=2}^{n^2} |e_j/d|_\infty$$

if we use (37) to estimate $|L_j^{(\infty)}(e_1/d, \ldots, e_{n^2}/d)|_\infty$. This optimal upper bound is achieved by the choice

$$L_j^{(\infty)} = y_j - y_1$$

for $j = 2, \ldots, n^2 - 1$ and $L_1^{(\infty)} = y_{n^2} - y_1$.

We note that $|e_{n^2}/d|_v$ is minimal and $|e_1/d|_v$ is maximal among the $e_j/d$ for each $v = 2, 3$. A similar argument to the above shows that by setting

$$L_j^{(v)} = y_j - y_{n^2}$$

for $j = 1, \ldots, n^2 - 1$ and $v = 2, 3$, we get the upper bound

$$\prod_{j=1}^{n^2-1} |L_j^{(v)}(e_1/d, \ldots, e_{n^2}/d)|_v \leq \prod_{j=1}^{n^2-1} |e_j/d|_v,$$

which is as good as can be if we use (37).

Multiplying our bounds, we get

$$\prod_{v \in S} \prod_{j=1}^{n^2-1} |L_j^{(v)}(e_1/d, \ldots, e_{n^2}/d)|_v \leq \prod_{j=2}^{n^2} |e_j/d|_\infty \cdot \prod_{v=2,3} \prod_{j=1}^{n^2-1} |e_j/d|_v.$$
We note that \(2 \nmid d\) and \(3 \nmid d\), for otherwise \(a\) and \(b\) would both be a power of the same prime, which is impossible by multiplicative inde-
pendence of \(a\) and \(b\). This shows that
\[
\prod_{v \in \mathcal{S}} |e_j/d|_v = \frac{1}{d}
\]
for each \(j\). Combining this observation with our previous bound we
obtain
\[
\prod_{v \in \mathcal{S}} \prod_{j=1}^{n^2-1} |L_j(v)(e_1/d, \ldots, e_{n^2}/d)|_v \leq \frac{a^n b^n}{d^{n^2-1}} \leq \max(a, b)^{2n - \varepsilon(n^2-1)}.
\]
Taking \(n = 3\varepsilon^{-1} + 1\), we get
\[
\prod_{v \in \mathcal{S}} \prod_{j=1}^{n^2-1} |L_j(v)(e_1/d, \ldots, e_{n^2}/d)|_v \leq \max(a, b)^{-n}.
\]
Since
\[
H(e_1/d, \ldots, e_{n^2}/d) \leq \max(a, b)^{2n},
\]
the subspace theorem applies.

This means that there is a finite collection \(\Lambda_1, \ldots, \Lambda_K \in V^*\) such that
\[
\Lambda_j(e_1/d, \ldots, e_{n^2}/d) = 0
\]
for some \(j\). These \(\Lambda_j\) depend only on \(n\) and hence only on \(\varepsilon\), and crucially not on \(a, b\). Now each \(\Lambda_j\) induces a linear form on \(\mathbb{Q}^{n^2}\), which we denote by the same symbol, and we clearly have
\[
\Lambda_j(e_1, \ldots, e_{n^2}) = 0.
\]
We apply Proposition 50 for each \(\Lambda_j\) and find that there is a constant \(C = C(\varepsilon)\) such that
\[
|e_i - e_j|_\infty |e_i - e_j|_2 |e_i - e_j|_3 \leq C
\]
for some \(i \neq j\). However, we have \(d|e_i - e_j\), hence \(d \leq C\), as required.

\begin{proof}[Proof of Proposition 50] We prove the proposition by induction on \(d\).
We first consider the case \(d = 2\). Let \(L = ax_1 + bx_2\), and let \(x_1, x_2 \in \mathcal{S}\)
be such that \(L(x_1, x_2) = 0\). We may assume without loss of generality that \(\gcd(x_1, x_2) = 1\). Indeed,
\[
|x_1 - x_2|_\infty |x_1 - x_2|_2 |x_1 - x_2|_3
\]
remains the same if we divide both \(x_1\) and \(x_2\) by their greatest common
divisor, which must be an element of \(\mathcal{S}\).

If \(\gcd(x_1, x_2) = 1\), then we must have \(x_2|a\) and \(x_1|b\), so there are only
finitely many choices of \(x_1\) and \(x_2\). Therefore, we can find a suitably
large \(C\) that works for all of these finitely many possibilities.
\end{proof}
Now we assume $d > 2$ and that the claim holds for linear forms in less than $d$ variables. We may assume that all coefficients of $L$ are non-zero, for otherwise the claim follows from the induction hypothesis.

Let $x_1, \ldots, x_d \in S$ with $L(x_1, \ldots, x_d) = 0$. We assume, as we may, that for each $v \in \{2, 3\}$, there is $x_j$ with $|x_j|_v = 1$. If this was not the case, we could divide each $x_j$ with their greatest common divisor.

For simplicity, we assume $|x_j|_\infty$ is maximal for $j = d$. We let $w \in \{2, 3\}$ be such that $|x_d|_w \leq |x_d|_\infty - 1/2$. We also assume for simplicity that $|x_1|_w = 1$. We can achieve that our simplifying assumptions hold by a suitable permutation of the coordinates. This changes the linear form $L$, but we can conclude the general statement of our proposition, by taking the maximum of the constants $C$ produced by our argument over all permutations of the coefficients of $L$.

If we take the linear forms $L_j^{(v)} = x_j$ for $v \in S$ and $j = 1, \ldots, d - 1$, then we get
\[
\prod_{v \in S} \prod_{j=1}^{d-1} |L_j^{(v)}(x_1, \ldots, x_{d-1})|_v = 1.
\]

Let $L = a_1 x_1 + \ldots + a_d x_d$. If we replace $L_1^{(w)}$ by $(a_1/a_d) x_1 + \ldots + (a_{d-1}/a_d) x_{d-1}$, then we get
\[
|L_1^{(w)}(x_1, \ldots, x_{d-1})|_w = |x_{d-1}|_w \leq |x_{d-1}|^{-1/2}
\]
instead of $L_1^{(w)}$. We also note that
\[
H(x_1, \ldots, x_{d-1}) \leq |x_d|_\infty,
\]
and hence the subspace theorem applies for this modified set of linear forms. Therefore, there is a finite set of linear forms depending only on $L$ such that at least one of them vanishes on $x_1, \ldots, x_{d-1}$. Now the claim follows by the induction hypothesis. □

7. The Catalan equation

This section is devoted to the following result of Tijdeman [22], which is an application of lower bounds for linear forms in logarithms.

**Theorem 51 (Tijdeman).** The solutions of the equation
\[
x^m - y^n = 1, \quad x, y, m, n \in \mathbb{Z}_{\geq 2}
\]
are bounded by an effective absolute constant.

The equation in the theorem is known as the Catalan equation, named after Eugén Catalan, who asked whether there are any pairs of positive proper powers whose difference is 1 beside 8 and 9.

Tijdeman’s constant has been made explicit by Langevin [14] and later by Mignotte [17]. The latter proved that the equation has no solutions if $m, n$ are primes and $\min(m, n) > 7.15 \times 10^{11}$. He also showed that the equation has no solution (other than $3^2 - 2^3 = 1$) if
\[
\min(m, n) < 10^7, \text{ which was later improved to } 3.2 \times 10^8 \text{ by Grantham and Wheeler.}
\]

Catalan’s problem was solved by Mihăilescu [18] along different lines. (He showed that there are no solutions other than \(3^2 - 2^3 = 1\).

On the other hand, we still do not know if the difference between consecutive proper powers tend to infinity.

Our exposition is based on Bugeaud’s book [9, Section 4.8].

*Partial proof of Theorem 51.* We first observe that it is enough to prove the theorem in the case when \(n\) and \(m\) are primes. Indeed, if \(x, y, m, n\) is a putative solution of the equation, then we may replace \(n\) and \(m\) by prime divisors of each and \(x\) and \(y\) by suitable powers of them.

We will prove the theorem only in the case when the following holds:

- \(n > m\),
- both \(n\) and \(m\) are odd primes,
- \(n > C\), for a suitably large absolute constant \(C\).

The case \(n < m\) can be dealt with trivial modifications. The case when at least one of \(m\) and \(n\) equals 2 requires a different proof, and so does the case, when \(m\) and \(n\) are both bounded by an absolute constant. This is why our proof is only partial.

The proof is based on lower bounds for linear forms in logarithms. To use such estimates, we need to find two integers that involve high powers, and which are close to each other. It is tempting to use \(x^m\) and \(y^n\). We can write

\[
|m \log x - n \log y| \leq C|x^m y^{-n} - 1| = Cy^{-n} = C \exp(-n \log y).
\]

On the other hand, we have

\[
|m \log x - n \log y| \geq \exp(-c \log x \log y \log n),
\]

by the best known lower bounds for linear forms in logarithms. Comparing our lower and upper bounds, we get

\[
n \log y \leq C \log x \log y \log n,
\]

which might look useful, but we need to do better. We want the bases of our powers to be smaller, and it would also help if those that are large were raised to the same power.

We can get better bounds using the following observations. First, we note that

\[
(x - 1) \frac{x^m - 1}{x - 1} = y^n,
\]

and the two factors on the left hand side are almost relatively prime. Indeed, we have

\[
\frac{x^m - 1}{x - 1} = \frac{(x - 1 + 1)^m - 1}{x - 1} \equiv m + \binom{m}{2} (x - 1) \mod (x - 1)^2.
\]
This shows that $m$ is the only prime that may divide both $(x^m - 1)/(x - 1)$ and $x - 1$. Moreover, if $m|(x - 1)$, then $m^2 \nmid (x^m - 1)/(x - 1)$. For this reason, we must have

$$x - 1 = \frac{u^n}{m^*},$$

where $u \geq 2$ is an integer dividing $y$ and $m^* \in \{1, m\}$. To see that $u = 1$ is not possible, we note that in that case we would have $x = 2$, and considering the factorization

$$(y + 1)\frac{y^n + 1}{y + 1} = 2^m,$$

we would have either $y + 1 = 1$ or $(y^n + 1)/(y + 1) = 1$, both of which are nonsense.

A similar argument gives

$$y + 1 = \frac{v^m}{n^*},$$

where $v \geq 2$ and $n^* \in \{1, n\}$.

Now we can try using the number $(u^n/m^*)^m$, which is close to $x^m$ and the number $(v^m/n^*)^n$, which is close to $y^n$ in our application of lower bounds on linear forms in logarithms. More precisely, we can write

$$\left|(u^n/m^*)^m - (v^m/n^*)^n\right| = \left|x^m(1 - 1/x)^m - y^n(1 + 1/y)^n\right| \leq Cy^n \cdot \frac{n}{y}$$

provided $y > n$, which we assume temporarily. (Note that $x > y$, since $n > m$, and hence $m/x \leq n/y$.)

Now we have

$$\left|(u^n/m^*)^m - (v^m/n^*)^n\right| \leq C \frac{n}{y},$$

which implies

$$|nm \log(u/v) - m \log(m^*) + n \log(n^*)| \leq C \frac{n}{y} \leq \exp(-c(\log y - \log n))$$

provided $y > Cn$ for a suitably large $C$, which we again assume temporarily. We compare this with the lower bound

$$|nm \log(u/v) - m \log(m^*) + n \log(n^*)| \geq \exp(-C \log(nm) \log v \log(m^*) \log(n^*))$$

and get

$$\log y \leq C \log v(\log n)^3.$$

Strengthening further our temporary hypothesis to $y > n^2$, we get

$$\log y = \log(v^m/n^* - 1) \geq cm \log v.$$

Comparing this with our previous bound, we obtain

$$m \leq C(\log n)^3.$$

Finally, we note that if our temporary hypothesis $y > n^2$ fails, then

$$v^m = (y + 1)n^* \leq 2n^3,$$
so we see that the conclusion

\[ m \leq C(\log n)^3 \]

holds always.

Now it remains to consider the case when \( m \) is much smaller than \( n \) and consequently, \( x \) is much larger than \( y \). In this situation, there is little to gain from replacing \( y^n \) by \((v^m/n^*)^n\), because \( \log y \) is already very small. If we apply our argument to \((u^n/m^*)^m\) and \( y^n \), then we can replace the error term \( n/y \) by the much smaller \( m/x \). More precisely, we can write

\[ \left| (u^n/m^*)^m - y^n \right| = \left| x^m (1 - 1/x)^m - y^n \right| \leq C x^m \cdot \frac{m}{x}. \]

Here, we used that

\[ x = \frac{u^n}{m^*} - 1 \geq m. \]

In fact, we have \( x \geq C m \) for an arbitrarily large constant \( C \), provided \( \min(x, y, m, n) \) is sufficiently large. This means that

\[ |n \log(u^n/y) - m \log(m^*)| \leq C \frac{m}{x}. \]

We note that \( y \leq v^m \leq (10u)^m \), hence we have

\[ |n \log(u^n/y) - m \log(m^*)| \geq \exp(-C m \log u \log m \log n). \]

Comparing this with the upper bound, and using

\[ x \geq u^n/(2m), \]

we get

\[ n \log u - 2 \log m \leq C m \log m \log n \log u, \]

which yields

\[ n/\log n \leq C m \log m. \]

Comparing this with our previous inequality between \( m \) and \( n \), we see that \( m \) and \( n \) must be bounded by an absolute constant. \( \square \)

\textbf{References}


