DIOPHANTINE ANALYSIS, MICHAELMAS 2024

PÉTER P. VARJÚ

Disclaimer, acknowledgement, etc

I prepared these notes primarily for the benefit of the students taking my lectures on Diophantine Analysis that I gave in the Michaelmas term, 2024 at the University of Cambridge. Others may find one of the standard textbooks on the subject more useful, from which I also borrowed extensively while writing these notes.

Masser's book [9] is a masterfully written and wholly entertaining account of the general ideas explored in this course, although the actual intersection of content is not very large. For a short and self-contained proof of Roth's theorem, we refer to Cassels [5, Chapter VI]. Bombieri and Gubler [2] give an excellent exposition of the full proof of the subspace theorem, discuss some of its applications and much more.

Waldschmidt [13] gives a very thorough account of the proofs of various bounds on linear forms in logarithms of algebraic numbers reaching (very nearly) the sate of the art. Bougeaud [3] focuses on applications, but also gives self-contained proofs of powerful bounds for linear forms in two logarithms. Baker [1] gives an excellent exposition of his original method.

Heights are discussed in [13, Chapter 3], [9, Chapter 14] and [2, Chapter 1].

All results in these notes are due to others. Any novelty must lie in the mistakes I have introduced unintentionally.

Please send comments to: pv270@dpmms.cam.ac.uk.

Examinable material

Non-examinable material is marked explicitly and is placed between vertical lines on both sides. Everything else in these notes are examinable, and the purpose of this section is to clarify what this means.

The exam will be set with the aim that the questions our reasonable and possible to solve by students who studied the examinable material. However, it is impossible to guarantee that all questions will be entirely unrelated to non-examinable material. Questions will contain unseen problem elements, and these can be anything related to the course as long as they are deemed reasonable assuming knowledge of the examinable material.

As an example, Roth's lemma (for polynomials of more than 2 variables) is stated in Section 3.5, which is marked non-examinable. (The 2 variable case discussed in Section 3.3 *is* examinable.) For this reason, a question asking to state Roth's lemma is not permissible on the exam. However, a question providing the statement of this lemma and asking for its proof or suggesting its use in a problem is allowed as long as the question is deemed reasonable in the context of the exam. This example was given to illustrate a principle and not with any particular potential question in mind. (The lecturer would find it challenging, if possible at all, to set a reasonable question of this kind.)

Non-examinable designation is limited to the parts explicitly marked that way and does not extend to other parts. For example, if these notes contained Theorem X whose proof relied on Lemma Y and the proof of Lemma Y was marked non-examinable, the proof of Theorem X assuming Lemma Y would remain examinable. While this issue does not seem to arise in the context of this course, it did in a previous one, which led to misunderstanding, disappointment and an appeal.

Material may be marked non-examinable broadly for two reasons. Either it relies on a prerequisite beyond what is assumed in the course, or that the material was not discussed in sufficient detail during the lectures. On the other hand, the parts are included in the notes because they aid understanding the material. For this reason non-examinable parts are recommended not to be ignored entirely.

1. Introduction

The course will 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 them.

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 α, there are infinitely many rational numbers p/q such that*

$$|\alpha - p/q| \leq \frac{1}{q^2}. \tag{1}$$

Proof. We fix an integer $N > 0$. We consider the image of the numbers $0, \alpha, 2\alpha, \ldots, N\alpha$ in $\mathbf{R}/\mathbf{Z}$. We identify $\mathbf{R}/\mathbf{Z}$ with $[0, 1]$, which we subdivide into 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 $\mathbf{R}/\mathbf{Z}$ 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 α 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 $d \in \mathbf{Z}_{\geq 1}$. For every real algebraic number α of degree d over $\mathbf{Q}$, there is constant $c = c(\alpha)$ such that*

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

for all rational $p/q \neq \alpha$.

Proof. Let $P(X) \in \mathbf{Z}[X]$ be the minimal polynomial of α. Let $p/q \in \mathbf{Q}$. 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 $\mathbf{Z}[X]$ being the minimal polynomial of α, and if $d = 1$, $\alpha \neq p/q$.

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

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

□

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 ≥ 1.

Liouville's theorem shows that for quadratic (i.e. algebraic of degree 2) irrationals, Dirichlet's theorem is best possible apart form 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

$$\min_{s=1,\ldots,d-1}\Big(s+\frac{d}{s+1}+\varepsilon\Big)<2\sqrt{d},$$

and by Dyson, Gelfond and Schneider 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 α 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 $P(X,Y) \in \mathbf{Z}[X,Y]$ be a homogeneous polynomial of degree at least 3 without repeated factors, and let $m \in \mathbf{Z}$ be an integer. Then the equation*

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

has only finitely many solutions in $x, y \in \mathbf{Z}$ with $\gcd(x,y) = 1$.

The assumption that P has no repeated factors means that the polynomial $P(X,1)$ has simple roots. This 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 $P(X,Y) \in \mathbf{R}[X,Y]$ be a homogeneous polynomial without repeated factors, and let $K \subset \mathbf{R}$ be a compact set. Suppose $P(X,Y)$ is not proportional to Y. Then there are constants $c = c(P,K)$ and $C = C(P)$ such that for all $x, y \in \mathbf{R}$ with $x/y \in K$ there is a root α of the polynomial $P(X,1)$ with*

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

Proof of Theorem 4. Now let $m \neq 0$. Then $P(x, y) = m$ and Lemma 5 implies

$$|\alpha - x/y| \leq C|y|^{-d}|P(x, y)| = C|m||y|^{-d}.$$

For some root α of $P(X, 1)$. When α is irrational, Roth's theorem (or any of its precursors discussed above) implies that there are only finitely many solutions. When α is rational, this follows by Liouville's theorem applied with $d = 1$. □

In fact, we could even replace m on the right hand side of the Thue equation by any polynomial (in $\mathbf{Z}[X, Y]$) of degree at most $d - 3$.

Proof of Lemma 5. Let

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

Since $P(X, Y)$ has no repeated factors, d' is d or $d - 1$. We note that

$$P(X, Y) = aY^{d-d'} \prod_{i=1}^{d'} (X - \alpha_i Y).$$

Suppose that α_1 is the one nearest to x/y. Then there are constants $c_0 = c_0(P)$ and $C_0 = C_0(P, K)$ such that $c_0 \leq |\alpha_i - x/y| \leq C_0$ for all $i = 2, \ldots, d'$. Indeed, $c_0 = \min \frac{|\alpha_i - \alpha_j|}{2}$ would work.

Then

$$|P(x, y)| = |a|y^d \prod_{i=1}^{d'} |x/y - \alpha_i| \geq y^d |x/y - \alpha_1| \cdot |a| c_0^{d'-1}.$$

Similarly

$$|P(x, y)| = |a|y^d \prod_{i=1}^{d'} |x/y - \alpha_i| \leq y^d |x/y - \alpha_1| \cdot |a| C_0^{d'-1}.$$

The claim follows. □

1.3. **The subspace theorem.** The subspace theorem is a far reaching generalization of Roth's theorem proved by Schmidt. For an integer vector $(x_1, \ldots, x_n) \in \mathbf{Z}^n$, 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, Archimedean version, Schmidt)**.** *Let $n \in \mathbf{Z}_{\geq 2}$, and let $L_1, \ldots, L_n$ be linearly independent linear forms in n variables with algebraic coefficients. Then for all $\varepsilon > 0$, the solutions of*

$$(2) \qquad \prod_{j=1}^{n} |L_j(x_1, \ldots, x_n)| \leq H(x_1, \ldots, x_n)^{-\varepsilon}, \qquad (x_1, \ldots, x_n) \in \mathbf{Z}^n$$

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

Observe that the volume of the set of points $(x_1, \ldots, x_n) \in \mathbf{R}^n$ defined by the inequalities $\prod_{j=1}^n |L_j(x_1, \ldots, x_n)| \le h^{-\varepsilon}$ is $|x_1|, \ldots, |x_n| \le h$ is less than $C(\log h)^{n-1} h^{-\varepsilon}$ for some constant C depending on the linear forms, which is (fast) decreasing as h grows. As a matter of intuition, we expect that the number of points in $\mathbf{Z}^d \cap A$ for some set $A \subset \mathbf{R}^d$ will be proportional to the volume of A. (But we can choose A, of course, to make this fail miserably.) On this basis, we do not expect to see many solutions of (2). The theorem shows that the solutions, if they exist, must be "special" in a certain geometric sense.

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

$$L_1(X_1, X_2) = X_1 - \alpha X_2, \qquad L_2(X_1, X_2) = X_2$$

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

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

By adjusting ε, we can eliminate the constant when q is sufficiently large, and we conclude that (p, q) is contained in a finite collection of proper subspaces, which depend only on α and ε. 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, $|\alpha - p/q| \ge q^{-2-\varepsilon}$ holds with finitely many exceptions, and hence $|\alpha - p/q| \ge Cq^{-2-\varepsilon}$ 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 $\mathrm{Ker}(L_j)$ 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 $\varepsilon \le 1$ by Dirichlet's theorem. In this case, the subspace $V = \{(X_1, X_2, 0)\}$ is among those that arise in the conclusion of the subspace theorem, but it is different from all $\mathrm{Ker}(L_j)$.
- A line (that is a 1 dimensional subspace of $\mathbf{Q}^n$) 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.

Schlickewei generalized the subspace theorem to allow p-adic absolute values. Similar generalizations of Roth's theorem have 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. Later in the course we will state a more general version. Bombieri and Gubler [2, Chapter 7] is a standard reference on this subject.

We introduce some terminology. The **places** $M_{\mathbf{Q}}$ of $\mathbf{Q}$ is the set of prime numbers together with the symbol ∞. For each place v, we endow $\mathbf{Q}$ 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 $p/q \leq 0$.

Note that

$$|xy|_v = |x|_v|y|_v, \qquad |x+y|_v \leq |x|_v + |y|_v$$

for all $x, y \in \mathbf{Q}$ and $v \in M_{\mathbf{Q}}$. Moreover, when v is finite, the triangle inequality improves to the ultrametric inequality

$$|x+y|_v \leq \max(|x|_v, |y|_v).$$

Theorem 7 (Subspace theorem, p-adic version, rational coefficients)**.** *Let $n \in \mathbf{Z}_{\geq 2}$. Let S be a finite finite set of places of $\mathbf{Q}$ containing ∞. For each $v \in S$, let $L_1^{(v)}, \ldots, L_n^{(v)}$ be linearly independent linear forms in n variables with rational coefficients. Then for all $\varepsilon > 0$, the solutions of*

$$\prod_{v\in S}\prod_{j=1}^{n} |L_j^{(v)}(x_1, \ldots, x_n)|_v \leq H(x_1, \ldots, x_n)^{-\varepsilon} \tag{3}$$

for $(x_1, \ldots, x_n) \in \mathbf{Z}^n$ lie in a finite union of proper subspaces of $\mathbf{Q}^n$.

Our remark about the finiteness of solutions contained in lines still stands. Now it is no longer true that the left hand side of (3) increases as we move along (integer points contained in) a line, but it will still be 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 (3) 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 8. *For all $\varepsilon > 0$, there is $c = c(\varepsilon)$ such that either*

$$|p2^k - q3^m| \geq c\frac{\max(2^k, 3^m)^{1-\varepsilon}}{\max(p, q)} \tag{4}$$

or $p2^k = q3^m$ for all $p, q, k, m \in \mathbf{Z}_{>0}$.

Here we wrote 2 and 3 for the sake of concreteness, but they could be replaced by any other integers.

Remark 9. Using the box principle, it is easy to see that this result is best possible apart from the $-\varepsilon$ in the exponent. Indeed, consider all numbers of the form $p2^k - q3^m$ for fixed k and m and $p, q \in \{0, \ldots, H\}$ for some fixed H. There are $(H+1)^2$ of these numbers and they all fall in the interval $[-3^m H, 2^k H]$, whose length is less than $2\max(2^k, 3^m)H$. It follows that there are two among these numbers that are of distance less than $2\max(2^k, 3^m)/H$. Taking the difference we find p, q not both 0 with $|p|, |q| \leq H$ such that

$$|p2^k - q3^m| \leq 2\frac{\max(2^k, 3^m)}{H} \leq 2\frac{\max(2^k, 3^m)}{\max(|p|, |q|)}.$$

Moreover, if we take $H < \max(2^k, 3^m)$, then $p2^k \neq q3^m$, because this would require $2^k|q$ and $3^m|p$.

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 p-adic subsapce theorem, the solutions of

$$\prod_{v \in S} \prod_{j=1}^{n} |L_j^{(v)}(X_1, X_2)|_v \leq H(X_1, X_2)^{-\varepsilon/2} \tag{5}$$

lie in a finite union of lines. As we discussed earlier, each of these lines may contain only finitely many solutions or one of the linear forms must vanish on the line. That is, there are only finitely many solutions not satisfying $X_1 = 0$, $X_2 = 0$ or $X_1 = X_2$.

Let $p, q, k, m \in \mathbf{Z}_{>0}$, and take $X_1 = p2^k$, $X_2 = q3^m$. 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 $H(p2^k, q3^m) \leq 3^{2m}$.

We have

$$\begin{aligned} |L_1^{(\infty)}(p2^k, q3^m)|_\infty &= p2^k, & |L_2^{(\infty)}(p2^k, q3^m)|_\infty &= |p2^k - q3^m| \\ |L_1^{(2)}(p2^k, q3^m)|_2 &\leq 2^{-k}, & |L_2^{(2)}(p2^k, q3^m)|_2 &\leq 1 \\ |L_1^{(3)}(p2^k, q3^m)|_3 &\leq 1, & |L_2^{(3)}(p2^k, q3^m)|_3 &\leq 3^{-m}. \end{aligned}$$

Assuming that (4) does not hold with $c = 1$, we get

$$\prod_{v\in S}\prod_{j=1}^{n}|L_j^{(v)}(p2^k, q3^m)|_v \leq p3^{-m}\cdot\frac{\max(2^k,3^m)^{1-\varepsilon}}{\max(p,q)} \leq 3^{-\varepsilon m} \leq H(p2^k, q3^m)^{-\varepsilon/2}.$$

That is, (5) holds for $x_1 = p2^k$, $x_2 = q3^m$. In light of the first half of the proof, there are only finitely many quadruples p, q, k, m such that (4) does not hold with $c = 1$ and $p2^k \neq q3^m$. 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. (Two integers are multiplicatively independent if we cannot raise each of them to some integer power and get the same integer.) 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 [11] as an application of Proposition 8.

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 $a \to \infty$*.*

Here 2 and 3 may be replaced by a pair of multiplicatively independent integers ≥ 2.

Proof. We show that for each $N \in \mathbf{Z}_{>0}$, there are only finitely many $a \in \mathbf{Z}_{>0}$ 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, α_I is the integer whose base 2 expansion is the sequence of those α_j for which $2^j = a^t$ for some $t \in I$. We define β_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 $2^{k_j} \ge a^{1-3^{-j-1}}$. We note that

$$\begin{aligned} p \le& a/2^{k_j} \le a^{3^{-j-1}}, \\ e_1 \le& 2a^{1-3^{-j}}. \end{aligned}$$

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

$$\begin{aligned} q \le& a^{3^{-j-1}}, \\ e_2 \le& 3a^{1-3^{-j}}. \end{aligned}$$

Observe that

$$|e_1 - e_2| \max(p, q)/2^{k_j} \le 3a^{1-3^{-j}} a^{3^{-j-1}} a^{-1+3^{-j-1}} = 3a^{-3^{-j}+2\cdot 3^{-j-1}}.$$

Therefore, we have

$$|p2^{k_j} - q3^{m_j}| = |e_1 - e_2| \le 3a^{-3^{-j}+2\cdot 3^{-j-1}} \frac{2^{k_j}}{\max(p, q)}.$$

We apply Proposition 8 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. □

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 we could have written any number there. (Well, impossible is maybe too strong a word here, but see Masser [9, 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. We 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 for 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 α 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 numbers 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, π or $2^{\sqrt{2}}$ would follow in this way. (However, to the best of my knowledge this has not been proved for π 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^α is transcendental if α is algebraic. This contains the transcendence of e (take $\alpha = 1$), and also that of π, 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 π), whose proof was completed by Weierstrass.

Theorem 11 (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 $\overline{\mathbf{Q}}$.*

Here and everywhere in these notes, $\overline{\mathbf{Q}}$ denotes the field of algebraic numbers.

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

Problem 12 (Hilbert's 7'th problem)**.** Is α^β always transcendental for algebraic $\alpha \neq 0, 1$ and irrational algebraic β?

The meaning of α^β 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 β is not an integer, then α^β is not uniquely determined, and the question is asked about any of the choices. On the other hand, in these notes, for $\alpha \in \mathbf{R}_{>0}$, we always mean by $\log \alpha$ its principal branch, that is $\log \alpha \in \mathbf{R}$ 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 13 (Gelfond, Schneider)**.** *Let* α_1, α_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* $\overline{\mathbf{Q}}$ *if and only if they are linearly independent over* $\mathbf{Q}$*.*

We prove this result later in the course, (with full details when $\alpha_1, \alpha_2 \in \mathbf{R}_{>0}$).

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 $\overline{\mathbf{Q}}$, then there is a number $\beta \in \overline{\mathbf{Q}}$ such that $\beta \log \alpha_1 = \log \alpha_2$. Exponentiating this, we get that $\alpha_1^\beta = \alpha_2$ is algebraic. So either $\beta \in \mathbf{Q}$ or $\alpha_1 = 1$. In both cases $\log \alpha_1$ and $\log \alpha_2$ are linearly dependent over $\mathbf{Q}$. (Here we used our convention $\log 1 = 0$.)

Now we prove the converse implication. Assume that $\alpha_1^\beta = \alpha_2$ is algebraic for some algebraic α_1 and β. 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 $\mathbf{Q}$. So either $\log \alpha_1 = \log \alpha_2 = 0$ and $\alpha_1 = 1$ or $\beta \in \mathbf{Q}$. In either case, the proof is complete. $\square$

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

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

field
$$\mathbf{Q}(\alpha_1, \ldots, \alpha_n, e^{\alpha_1}, \ldots, e^{\alpha_n})$$
is at least n.

This conjecture strengthens the Lindemann Weierstrass theorem in that for $\mathbf{Q}$ linearly independent algebraic numbers $\alpha_1, \ldots, \alpha_n$ it claims not only the $\overline{\mathbf{Q}}$ 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 of the Lindemann Weierstrass theorem.

On the other hand, Schanuel's conjecture also implies a strengthening of Baker's theorem about the algebraic independence of ($\mathbf{Q}$ linearly independent) logarithms of algebraic numbers. This strengthening is wide open. Even the following simple case is not known. Let $\log \alpha_1, \ldots, \log \alpha_4$ be four non-zero logarithms of algebraic numbers such that $\log \alpha_1 / \log \alpha_2, \log \alpha_1 / \log \alpha_3 \notin \mathbf{Q}$. 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 (prominently including Baker himself) 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 [3] or the slightly older book of Waldschmidt [13].

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

$$|b_1 \log a_1 + \ldots + b_n \log a_n|. \tag{6}$$

This is closely related to the quantity

$$|a_1^{b_1} \cdots a_n^{b_n} - 1|. \tag{7}$$

By expanding exp around 0, it is easy to see that when (6) is small, then (7) 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 (6) is close to $2\pi i k$ for some $k \in \mathbf{Z}$ when (7) is small. We will need to pay attention to this issue.

We observe that (7) is a rational number with denominator at most $A_1^{|b_1|}\cdots A_n^{|b_n|}$, where A_j is the maximum of the numerator and the denominator of a_j. 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.

We also need some notation. For an algebraic number $\alpha \in \overline{\mathbf{Q}}$, we write $f_\alpha \in \mathbf{Z}[X]$ for its minimal polynomial. For a polynomial $f \in \mathbf{C}[X]$, we write $H(f)$ for the maximal absolute value of its coefficients, and call it the height of f.

First we state a general result for inhomogeneous forms.

Theorem 16. *Let $n \in \mathbf{Z}_{\geq 1}$. Let $\alpha_1, \ldots, \alpha_n \in \overline{\mathbf{Q}}_{\neq 0}$ and let $\log \alpha_j$ be any choice of the logarithm of α_j for each $j = 1, \ldots, n$. Let $\beta_0, \ldots, \beta_n \in \overline{\mathbf{Q}}$ and let*

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

Let

$$A_j := \max(H(f_{\alpha_j}), \exp(|\log \alpha_j|), 10)$$

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

$$B := \max(H(f_{\beta_j}), \log(A_i)).$$

Then there exists an effective constant C depending only on n and the degree of $\mathbf{Q}(\alpha_1, \ldots, \alpha_n, \beta_0, \ldots, \beta_n)$ 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 17. *Let $n \in \mathbf{Z}_{\geq 1}$. Let $\alpha_1, \ldots, \alpha_n \in \overline{\mathbf{Q}}_{\neq 0}$ and let $\log \alpha_j$ be any choice of the logarithm of α_j. Let $b_1, \ldots, b_n \in \mathbf{Z}$ and let*

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

Let

$$A_j := \max(H(f_{\alpha_j}), \exp(|\log \alpha_j|), 10)$$

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

$$B^* := \max\Big(\frac{|b_1|}{\log A_n}, \ldots, \cdot \frac{|b_{n-1}|}{\log A_n}, |b_n|, 10\Big).$$

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

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

Both of the above theorems follow from [13, Theorem 9.1] if we set the parameter E there to be 10, say. 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 $\mathbf{Q}$ linearly independent. However, it is useful in applications not to restrict ourselves to that setting.
- Theorems 16 and 17 dramatically improve 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 [13, Conjecture 14.25].
- The improvement in Theorem 17 may seem very minor. In some applications, one has that $\alpha_1, \ldots, \alpha_n$ are fixed numbers, $b_n = 1$ and $\log H(f_{\alpha_n})$ is roughly of the same size as $b_1, \ldots, b_{n-1}$. In such cases, the quantity B^* offers an improvement over B, which can be significant.
- In the special case when the α_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 17 if n and the degree of $\mathbf{Q}(\alpha_1, \ldots, \alpha_n)$ 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 8.

Proposition 18. *There is an effective absolute constant C such that the following holds. For all $p, q, k, m \in \mathbf{Z}_{>0}$, we have*

$$|p2^k - q3^m| \geq \frac{\max(2^k, 3^m)}{\max(p, q, 10)^{-C \log(\max(k,m)/\log \max(p,q,10)+10)}}$$

or $p2^k = q3^m$.

Before the proof we make some comments.

- If we apply the weaker bounds in Theorem 16, we get the same result with the exponent $-C \log \max(k, m)$ in place of $-C \log(\max(k, m)/\log \max(p, q, 10))$.
- In addition to being effective, this result improves on Proposition 8 when $\max(p, q) < \max(2^k, 3^m)^{o(1)}$.

- In particular, we get
$$|2^k - 3^m| \geq \frac{\max(2^k, 3^m)}{\max(k,m)^{-C}}$$
for some C instead of
$$|2^k - 3^m| \geq c \max(2^k, 3^m)^{1-\varepsilon}.$$
- On the other hand, the scope of the p-adic subspace theorem is wider. For example, it is easy to give lower bounds for quantities like
$$p_1 2^{k_1} + p_2 3^{k_2} + p_3 5^{k_3}$$
using the method of proof of Proposition 8. However, this seems beyond reach using linear forms in logarithms.

Proof. We write $A = \max(p, q, 10)$ and assume for simplicity that $3^m > 2^k$. Suppose further that $p2^k \neq q3^m$. We apply Theorem 17 for
$$\Lambda = k \log 2 - m \log 3 + 1 \cdot \log(p/q)$$
and get
$$|\exp \Lambda - 1| = \Big|\frac{p}{q} 2^k 3^{-m} - 1\Big| \geq \exp\Big(- C \log A \log\Big(\frac{\max(k,m)}{\log A} + 10\Big)\Big).$$
Here we used an expansion of exp around 0 and that $\Lambda \in \mathbf{R}$.

Now we multiply the left hand side by $q3^m$ and the right by 3^m and get
$$|p2^k - q3^m| \geq \frac{3^m}{A^{-C \log(\max(k,m)/\log A + 10)}}.$$
This is precisely what we wanted to prove, since $3^m = \max(2^k, 3^m)$ and $A = \max(p, q, 10)$. □

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 19 (Stewart [12])**.** *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 $a \in \mathbf{Z}_{>2}$.

Again, here 2 and 3 may be replaced by a pair of multiplicatively independent integers ≥ 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 $a \in \mathbf{Z}_{>0}$ with a suitable choice of C, which will be made later.

We recall some notation from the proof of Theorem 10. We write $\alpha_k\alpha_{k-1}\dots\alpha_0$ and $\beta_m\beta_{m-1}\dots\beta_0$ 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, α_I is the integer whose base 2 expansion is the sequence of those α_j for which $2^j = a^t$ for some $t \in I$. We define β_I using the base 3 expansion of a in a similar manner.

By Proposition 18, there is some C_0 such that

$$|p2^{k'} - q3^{m'}| \ge \frac{\max(2^{k'}, 3^{m'})}{\max(p,q,2)^{C_0 \log k}} \tag{8}$$

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

By our assumption on a, there is some

$$j \in \{1, \dots, \frac{\log\log a}{\log\log\log a + C}\}$$

such that

$$\alpha_{(1-K^{j+1}/\log a, 1-K^j/\log a]} = \beta_{(1-K^{j+1}/\log a, 1-K^j/\log a]} = 0.$$

By setting C sufficiently large, which can be done independently of a, we can ensure that all of these intervals are contained in $[0,1]$. (Note that k is approximately $\log_2 a$.) 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 $a = p2^{k_j} + e_1$, where $p = \alpha_{[1-K^j/\log a, 1]}$, $e_1 = \alpha_{(0,1-K^{j+1}/\log a]}$ and k_j is the smallest integer so that $2^{k_j} \ge a^{1-K^j/\log a}$. We note that

$$\begin{aligned} p \le& a/2^{k_j} \le a^{K^j/\log a} \\ e_1 \le& 2a^{1-K^{j+1}/\log a}. \end{aligned}$$

The important thing for us, as we will see below, is that $e_1 < 2^{k_j}/p^{K-2}$. In a similar manner, we can write $a = q3^{m_j} + e_2$ with

$$\begin{aligned} q \le& a^{K^j/\log a} \\ e_2 \le& 3a^{1-K^{j+1}/\log a}. \end{aligned}$$

Therefore, we have

$$|p2^{k_j} - q3^{m_j}| = |e_1 - e_2| \le \frac{2^{k_j}}{\max(p,q)^{K-2}},$$

which contradicts (8). $\square$

2. Heights

We have already encountered the quantity $H(f_\alpha)$, the maximal absolute value of the coefficients of the minimal polynomial of $\alpha \in \overline{\mathbf{Q}}$. 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.

References are [13, Chapter 3], [9, Chapter 14] and [2, Chapter 1].

2.1. Mahler measure.

Let

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

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

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

When α and β are algebraic integers $[\mathbf{Q}(\alpha\beta):\mathbf{Q}] = [\mathbf{Q}(\alpha+\beta):\mathbf{Q}] = [\mathbf{Q}(\alpha):\mathbf{Q}][\mathbf{Q}(\beta):\mathbf{Q}]$, 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 α and β. Indeed, we write $\alpha_1,\ldots,\alpha_{d_1}$ for the Galois conjugates of α including itself and $\beta_1,\ldots,\beta_{d_2}$ for the Galois conjugates of β again including itself. Then the conjugates of $\alpha\beta$ (respectively $\alpha+\beta$) are $a_i\beta_j$ (respectively $\alpha_i+\beta_j$) for $i=1,\ldots,d_1$ and $j=1,\ldots,d_2$, and these are also all algebraic integers. We can write

$$\begin{aligned}
M(f_{\alpha\beta}) =& \prod_{i=1}^{d_1}\prod_{j=1}^{d_2} \max(1, |\alpha_i\beta_j|)\\
\le& \prod_{i=1}^{d_1}\prod_{j=1}^{d_2} \max(1, |\alpha_i|)\max(1, |\beta_j|)\\
=& M(f_\alpha)^{d_2} M(f_\beta)^{d_1}
\end{aligned}$$

and

$$\begin{aligned}
M(f_{\alpha+\beta}) =& \prod_{i=1}^{d_1}\prod_{j=1}^{d_2} \max(1, |\alpha_i+\beta_j|)\\
\le& \prod_{i=1}^{d_1}\prod_{j=1}^{d_2} 2\max(1, |\alpha_i|)\max(1, |\beta_j|)\\
\le& 2^{d_1d_2} M(f_\alpha)^{d_2} M(f_\beta)^{d_1}.
\end{aligned}$$

In the next section we will introduce the notion of height $H(\alpha)$ of an algebraic number α. We could define it as

$$H(\alpha) = M(f_\alpha)^{1/d}, \tag{9}$$

where d is the degree of α. However, it will be easier for us to give an alternative definition instead, and prove (9) as a proposition.

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

$$\log M(P) = \int_0^1 \log |P(e^{2\pi it})| dt. \tag{10}$$

for the Mahler measure, which follows easily from Jensen's formula in complex analysis. (If you do not know this result, try to convince yourself that it is enough to check it for polynomials of degree ≤ 1.)

Theorem 20. *Let $P \in \mathbf{C}[x]$ of degree d. Then*

$$2^{-d} H(P) \leq M(P) \leq (d+1)H(P).$$

Proof. To prove the upper bound, we use (10). We can write

$$\begin{aligned}\log M(P) &\leq \int_0^1 \log |P(e^{2\pi it})| dt \\ &\leq \int_0^1 \log((d+1)H(P))dt \\ &= \log((d+1)H(P)),\end{aligned}$$

and the claim follows by exponentiation.

To prove the lower bound, we can estimate the coefficients by

$$\frac{|a_k|}{|a_d|} \leq \sum_{\{j_1,\dots j_k\}\subset\{1,\dots,d\}} |\alpha_{j_1}| \cdots |\alpha_{j_k}| \leq 2^d \prod_{j=1}^d \max(1, |\alpha_j|).$$

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. □

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. **Height.** In this section, we introduce the notion of height in a way that does not involve 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 α into $\mathbf{C}$ but also into p-adic fields.

We introduce the required notation and terminology. Let K be a number field, and let $\mathcal{O}_K$ be its ring of integers. An absolute value on K is a function $K \to \mathbf{R}_{\geq 0}$, $x \mapsto |x|$ that satisfies $|xy| = |x| \cdot |y|$, $|x + y| \leq |x| + |y|$ for all $x, y \in K$, which is not identically 0. We have the following examples.

- The trivial absolute value: $|x| = 1$ for all $x \neq 0$.
- Let $\sigma : K \to \mathbf{C}$ be an embedding. Then $|x|_\sigma = |\sigma(x)|$ is an absolute value, where $|\cdot|$ denotes the ordinary absolute value on $\mathbf{C}$.
- Let $P \subset \mathcal{O}_K$ be a non-zero prime ideal lying over the rational prime p. That is, p is the unique rational prime with $p \in P$. For $\alpha \in \mathcal{O}_K$, we define $\operatorname{ord}_P(\alpha)$ to be the largest integer m such that $P^m | \alpha\mathcal{O}_K$. We set $\operatorname{ord}_P(0) = \infty$. We extend this to K by $\operatorname{ord}_P(\alpha/\beta) = \operatorname{ord}_P(\alpha) - \operatorname{ord}_P(\beta)$. Write e_P for the ramification index of P, that is $\operatorname{ord}_P(p)$. Define $|x|_P = p^{-\operatorname{ord}_P(\alpha)/e_P}$ for $x \in K$.

  These absolute values satisfy a stronger form of the triangle inequality: $|x + y|_P \leq \max(|x|_P, |y_P|)$, which is called the ultrametric inequality.

By a theorem of Ostrowski all absolute values of K are of the form $|\cdot|^\alpha$, where $|\cdot|$ is one of the above examples and $\alpha \in \mathbf{R}_{>0}$. (These are not necessarily absolute values for all choices of α and $|\cdot|$.) The above normalization was chosen to coincide with the familiar absolute values when they are restricted to $\mathbf{Q}$. That is, we have $|x|_\sigma = |x|_\infty$ and $|x|_P = |x|_p$ for all $x \in \mathbf{Q}$, all embeddings $\sigma : K \to \mathbf{C}$ and all prime ideals $P \subset \mathcal{O}_K$, where p is the rational prime that lies below P.

We denote by M_K the places of K, which comprises all prime ideals of $\mathcal{O}_K$, all embeddings of K into $\mathbf{R}$ and one from each conjugate pair of complex embeddings. For $v \in M_K$, $|\cdot|_v$ will denote the absolute value defined above. We write $M_{K,f}$ for the finite places (i.e. the prime ideals) and $M_{K,\infty}$ for the infinite places (i.e. the embeddings).

For $v \in M_{K,\infty}$, we write $d_v = 1$ if v is a real embedding and $d_v = 2$ if it is complex. Let $v \in M_{K,f}$ be a finite place, that is a prime ideal. Let p be the rational prime below v, and recall the ramification index e_v. The inertial degree of v is $f_v = [\mathcal{O}_K/v : \mathbf{Z}/p\mathbf{Z}]$. We write

$$d_v = e_v f_v.$$

Recall from the theory of number fields that summing over all prime ideals lying over a given rational prime, we have

$$\sum_{v|p} d_v = [K : \mathbf{Q}].$$

Remark 21 (Non-examinable)**.** We do not need to know this, but it holds that $d_v = [K_v : \mathbf{Q}_p]$.

We made our normalization conventions so that the following handy product formula holds.

Lemma 22. *For all $\alpha \in K$, we have*
$$\prod_v |\alpha|_v^{d_v} = 1.$$

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

Proof. Consider the case $\alpha \in \mathcal{O}_K$ first. Note
$$\alpha\mathcal{O}_K = \prod_{v \in M_{K,f}} v^{\mathrm{ord}_v(\alpha)}.$$
Recall $|\alpha|_v = p^{-\mathrm{ord}_v(\alpha)/e_v} = p^{-\mathrm{ord}_v(\alpha) f_v/d_v}$.

Observe that
$$N(\alpha\mathcal{O}_K) = \prod_{v \in M_{K,f}} N(v)^{\mathrm{ord}_v(\alpha)} = \prod_{v \in M_{K,f}} p^{f_v \,\mathrm{ord}_v(\alpha)} = \prod_{v \in M_{K,f}} |\alpha|_v^{-d_v}.$$
On the other hand
$$N(\alpha\mathcal{O}_K) = |N_{K|\mathbf{Q}}(\alpha)| = \prod_{v \in M_{K,\infty}} |v(\alpha)|^{d_v}.$$
Dividing these two formulas gives the claim. For general α, the claim follows by multiplicativity of the absolute values. □

We define the **absolute height** or **Weil height** of $\alpha \in K$ by
$$H(\alpha) = \prod_v \max(1, |\alpha|_v)^{d_v/[K:\mathbf{Q}]}.$$
In the literature, the notation $h(\alpha) = \log H(\alpha)$ is also commonly used and it is called the logarithmic (absolute/Weil) height α.

Remark 23. The above notion of height is standard in the literature. However, the normalization conventions involving absolute values and places are not. Our choices are equivalent to those in [13, Chapter 3], but e.g. the conventions in [2, Chapter 1] and [9, Chapter 14] differ from ours and from each other. This impacts some of the formulas, including the one for the definition of height and the product formula.

Remark 24 (Non-examinable)**.** Let p be a rational prime. There is a field extension $\mathbf{C}_p$ of the p-adic numbers $\mathbf{Q}_p$ endowed with an absolute value $|\cdot|_p$ extending the p-adic absolute value on $\mathbf{Q}_p$ such that $\mathbf{C}_p$ is algebraicly closed and complete with respect to $|\cdot|_p$. We may think about this as an analogue of the field of complex numbers, which we now denote by $\mathbf{C}_\infty$.

Let α be an algebraic number, and write $f_\alpha \in \mathbf{Z}[x]$ for the minimal polynomial of α. Let $K = \mathbf{Q}(\alpha)$. It turns out that f_α has

$\deg f_\alpha$ distinct roots in $\mathbf{C}_p$ and their $|\cdot|_p$ absolute values are precisely the numbers
$$\{|\alpha|_v : v \in M_K, v|p\}.$$
Moreover, $|\alpha|_v$ is the $|\cdot|_p$ absolute value of d_v distinct roots of f_v. This allows us to write
$$H(\alpha) = \prod_{v\in M_{\mathbf{Q}}} \prod_{x\in\mathbf{C}_v : f_\alpha(x)=0} \max(1, |x|_v)^{1/[K:\mathbf{Q}]}.$$

Proposition 25. *The above definition of $H(\alpha)$ is independent of the choice of the number field K containing α.*

Proof. Let $L|K$ be a field extension. Let $w \in M_L$ and $v \in M_K$. We write $w|v$ if either both are infinite, that is, embeddings, and the restriction of w to K is v or $\overline{v}$, or if both are finite, that is, prime ideals, and $w|vO_L$.

The proposition will follow immediately if we show the following two statements.

Claim 1. For all $\alpha \in K$ and places $w|v$, we have $|\alpha|_w = |\alpha|_v$.

Claim 2. For all $v \in M_K$, we have
$$\sum_{w\in M_L : w|v} d_w = [L : K]d_v = \frac{[L : \mathbf{Q}]d_v}{[K : \mathbf{Q}]}.$$

We first prove Claim 1. If w and v are infinite places, then $|\alpha|_w = |w(\alpha)| = |v(\alpha)| = |\alpha|_v$ and the claim follows. If w and v are finite, then they lie over the same rational prime p. Since v is the only prime in $\mathcal{O}_K$ such that $w|v\mathcal{O}_L$, it follows that $\mathrm{ord}_w(v\mathcal{O}_L) = e_w/e_v$. Moreover, $\mathrm{ord}_w(\alpha) = (e_w/e_v)\,\mathrm{ord}_v(\alpha)$ for all $\alpha \in \mathcal{O}_K$. Therefore,
$$|\alpha|_w = p^{-\mathrm{ord}_w(\alpha)/e_w} = p^{-\mathrm{ord}_v(\alpha)/e_v} = |\alpha|_v.$$
The claim is proved.

We turn to Claim 2. First we consider the case that v is a real embedding. Then the sum $\sum_{w\in M_L:w|v} d_w$ contains a term for each real embedding of L extending v and one term for each pair of complex embeddings of L extending v. For the former kind of terms $d_w = 1$ for the latter kind $d_w = 2$. Therefore, $\sum_{w\in M_L:w|v} d_w$ is the number of embeddings of L into $\mathbf{C}$ (real or complex) that extend v. It is a fact of Galois theory that the number of these is $[L : K]$, as needed.

Next we consider the case that v is a complex embedding. Then $\{w \in M_L : w|v\}$ consists of embeddings of L that extend either v or $\overline{v}$, and it contains one from each pair of such complex conjugate embeddings. It follows that the number of terms equals the number of embeddings extending v, which is $[L : K]$. Now $d_w = 2 = d_v$ for all w, so the claim follows again.

Finally, suppose that v is a finite place. As we noted in the proof of Claim 1, the factorization of $v\mathcal{O}_L$ is

$$v\mathcal{O}_L = \prod_{w:w|v} w^{e_w/e_v}.$$

Since $N(w) = p^{f_w}$, we have

$$N(v\mathcal{O}_L) = p^{\sum_{w:w|v} d_w/e_v}.$$

On the other hand

$$N(v\mathcal{O}_L) = N(v)^{[L:K]} = p^{f_v[L:K]} = p^{[L:K]d_v/e_v}.$$

Comparing the exponents in the above two identities, the claim follows. □

Proposition 26. *We have*

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

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

$$2^{-d}H(f_\alpha) \leq H(\alpha)^d \leq (d+1)H(f_\alpha),$$

where $d = [\mathbf{Q}(\alpha) : \mathbf{Q}]$

Proof. Let K be a number field containing α. Comparing the definitions, it is enough to show that

$$\prod_{v\in M_{K,f}} \max(1, |\alpha|_v)^{d_v} = |a_d|^{[K:\mathbf{Q}(\alpha)]}, \tag{11}$$

where a_d is the leading coefficient of the minimal polynomial of α. Strictly speaking, the sufficiency of this can be seen from the definitions for $K = \mathbf{Q}(\alpha)$, but the previous proposition (and its proof) shows that we can take any number field K containing α. We take K to be the splitting field of f_α.

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 $v \in M_{K,f}$, we have

$$|Q_1Q_2|_v = |Q_1|_v \cdot |Q_2|_v.$$

Now we apply this to the factorization of the minimal polynomial of α. By definition of the minimal polynomial, we have

$$|a_d(x-\alpha_1)\cdots(x-\alpha_d)|_v = 1$$

for all $v \in M_{K,f}$. Therefore,

$$\prod_{v\in M_{K,f}} |a_d|_v^{d_v} \cdot \prod_{j=1}^{d} \prod_{v\in M_{K,f}} \max(1, |\alpha_j|_v)^{d_v} = 1.$$

Now we use that each α_j can be mapped to α by an automorphism of K, which permutes the places $M_{K,f}$ keeping d_v invariant. This gives

$$\prod_{v\in M_{K,f}} \max(1,|\alpha_j|_v)^{d_v} = \prod_{v\in M_{K,f}} \max(1,|\alpha|_v)^{d_v}$$

for all j. The product formula for a_d gives

$$\prod_{v\in M_{K,f}} |a_d|_v^{d_v} = \prod_{v\in M_{K,\infty}} |a_d|_v^{-d_v}.$$

We get

$$\Big(\prod_{v\in M_{K,f}} \max(1,|\alpha|_v)^{d_v}\Big)^d = \prod_{v\in M_{K,\infty}} |a_d|_v^{d_v}.$$

This implies (11), because $|a_d|_v = |a_d|$ for each infinite place, and

$$\sum_{v\in M_{K,\infty}} d_v = [K:\mathbf{Q}] = [K:\mathbf{Q}(\alpha)]\cdot d.$$

□

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

Lemma 27. *For all $\alpha\in K$ and $k\in\mathbf{Z}$ 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 $k=-1$. □

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. This is called the length of P.

Proposition 28. *Let $k\in\mathbf{Z}_{\geq 1}$, $n_1,\ldots,n_k\in\mathbf{Z}_{\geq 0}$, and let $P,Q\in\mathbf{Z}[x_1,\ldots,x_k]$ be two polynomials that are of degree at most n_j in the variable x_j for each j. Let $\alpha_1,\ldots,\alpha_k\in\overline{\mathbf{Q}}$. Then*

$$H\Big(\frac{P(\alpha_1,\ldots,\alpha_k)}{Q(\alpha_1,\ldots,\alpha_k)}\Big) \leq \max(\mathcal{L}(P),\mathcal{L}(Q))\prod_{j=1}^k H(\alpha_j)^{n_j}$$

Before giving the proof, we note the following two special cases of interest

$$H(\alpha_1\alpha_2)\leq H(\alpha_1)H(\alpha_2), \qquad H(\alpha_1+\alpha_2)\leq 2H(\alpha_1)H(\alpha_2).$$

Proof. Let K be a number field containing all α_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 $\mathbf{Z}$ whose $|\cdot|_v$ absolute value are at most 1, we can disregard them in our calculation. We can thus write

$$|P(\alpha_1,\ldots,\alpha_k)|_v \le \max_{m_j=0,\ldots,n_j:j=1,\ldots,k} \Big|\prod_{j=1}^k \alpha_j^{m_j}\Big|_v = \prod_{j=1}^k \max(1,|\alpha_j|_v)^{n_j}.$$

For an infinte place v, we use the triangle inequality for the monomials and get

$$\begin{aligned}|P(\alpha_1,\ldots,\alpha_k)|_v \le& \mathcal{L}(P) \max_{m_j=0,\ldots,n_j,j=1,\ldots,k} \Big|\prod_{j=1}^k \alpha_j^{m_j}\Big|_v\\ =&\mathcal{L}(P)\prod_{j=1}^k \max(1,|\alpha_j|_v)^{n_j}.\end{aligned}$$

We can write

$$H\Big(\frac{P(\alpha_1,\ldots,\alpha_k)}{Q(\alpha_1,\ldots,\alpha_k)}\Big)^{[K:\mathbf{Q}]} = \prod_v \max\Big(1,\frac{|P(\alpha_1,\ldots,\alpha_k)|_v}{|Q(\alpha_1,\ldots,\alpha_k)|_v}\Big)^{d_v}.$$

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

$$H\Big(\frac{P(\alpha_1,\ldots,\alpha_k)}{Q(\alpha_1,\ldots,\alpha_k)}\Big)^{[K:\mathbf{Q}]} = \prod_v \max\Big(|Q(\alpha_1,\ldots,\alpha_k)|_v,|P(\alpha_1,\ldots,\alpha_k)|_v\Big)^{d_v}.$$

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 $[K:\mathbf{Q}]$ infinite places taking the d_v multiplicities into account, we get

$$\begin{aligned}H\Big(\frac{P(\alpha_1,\ldots,\alpha_k)}{Q(\alpha_1,\ldots,\alpha_k)}\Big)^{[K:\mathbf{Q}]} &\le \max(\mathcal{L}(P),\mathcal{L}(Q))^{[K:\mathbf{Q}]}\prod_v\prod_{j=1}^k \max(1,|\alpha_j|_v)^{n_jd_v}\\ &= \max(\mathcal{L}(P),\mathcal{L}(Q))^{[K:\mathbf{Q}]}\prod_{j=1}^k H(\alpha_j)^{n_j\cdot[K:\mathbf{Q}]},\end{aligned}$$

which proves the claim. □

The following lemma gives lower and upper bounds for the absolute value of an algebraic number in terms of its height and degree, which are sharp if we do not have any other information about the number. This is referred to as the trivial bound or the Liouville bound.

Lemma 29. *For all $\alpha\in\overline{\mathbf{Q}}$, we have*

$$H(\alpha)^{-d}\le|\alpha|\le H(\alpha)^d,$$

where $d=\deg\alpha$.

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

3. Thue–Siegel–Dyson...

The purpose of this section is to prove the following theorem.

Theorem 30. *Let α be an irrational real algebraic number. Then for all $\varepsilon > 0$, there is a constant $c = c(\varepsilon, \alpha)$ such that*

$$|\alpha - p/q| \geq \frac{c}{q^{\sqrt{2d}+\varepsilon}}$$

for all rational numbers p/q.

At the end of the section, we outline how the proof may be improved to yield Roth's $2+\varepsilon$ exponent in place of $\sqrt{2d}+\varepsilon$. A full self-contained exposition of that can be found in the book of Cassels [5, 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 $p/q \in \mathbf{Q}$ with $|\alpha - p/q| < q^{-\sqrt{2d}-\varepsilon}$ and select a suitable pair $p_1/q_1, p_2/q_2$ among them.
(2) Find a suitable polynomial $P(X_1, X_2) \in \mathbf{Z}[X_1, X_2]$ that vanishes at the point (α, α) to high order.
(3) Give a lower bound for $|P(p_1/q_1, p_2/q_2)|$.
(4) Give an upper bound for $|P(p_1/q_1, p_2/q_2)|$.
(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 using a bivariate polynomial P in place of the minimal polynomial in α in a single variable. 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 $P \in \mathbf{Z}[X]$ of degree n, then it can have a zero of multiplicity at most n/d at α, where $d = [\mathbf{Q}(\alpha) : \mathbf{Q}]$. 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 α. 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 α, 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 just a single exceptionally close approximation to α existed and we were able to find it. 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 $|P(p_1/q_1, p_2/q_2)|$ 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. We will return to this point later. Once we show non-vanishing, we easily obtain the lower bound

$$|P(p_1/q_1, p_2/q_2)| \geq q_1^{-n_1} q_2^{-n_2}, \tag{12}$$

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 $q_1^{n_1} q_2^{n_2}$.

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

$$P(p_1/q_1, p_2/q_2) = \sum_{j_1, j_2} P_{j_1,j_2}(\alpha, \alpha)(\alpha - p_1/q_1)^{j_1}(\alpha - p_2/q_2)^{j_2}, \tag{13}$$

where

$$P_{j_1,j_2} = \frac{1}{j_1! j_2!} \cdot \frac{\partial^{j_1+j_2}}{\partial x_1^{j_1} \partial x_2^{j_2}} P.$$

For the purposes of this discussion, we will ignore the size of the coefficients $P_{j_1,j_2}(\alpha, \alpha)$ and also we will just look at largest non-zero term in (13). Later on we will need to estimate this more precisely, which will also require an estimate on the coefficients of P.

We observe that

$$\begin{aligned} |(\alpha - p_1/q_1)^{j_1}(\alpha - p_2/q_2)^{j_2}| \leq & q_1^{-(\sqrt{2d}+\varepsilon)j_1} q_2^{-(\sqrt{2d}+\varepsilon)j_2} \\ = & \exp(-(\sqrt{2d} + \varepsilon)(j_1 \log q_1 + j_2 \log q_2)). \end{aligned}$$

This motivates us to introduce the notion of the index of a function $F : \mathbf{R}^2 \to \mathbf{R}$ at a point (β_1, β_2) with respect to weights w_1, w_2 by

$$I_F(\beta_1, \beta_2; w_1, w_2) = \min_{j_1, j_2} \Big\{ j_1 w_1 + j_2 w_2 : F_{j_1,j_2}(\beta_1, \beta_2) \neq 0 \Big\}.$$

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 (13) in the form

$$\exp(-(\sqrt{2d} + \varepsilon) I_P(\alpha, \alpha; \log q_1, \log q_2)). \tag{14}$$

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 (12) is larger than (14). To this end, we need to maximize the index of P at (α, α) with respect to the weights $\log q_1, \log q_2$.

We fix some numbers n_1, n_2 and look for P in the form

$$P = \sum_{j_1=0}^{n_1} \sum_{j_2=0}^{n_2} a_{j_1,j_2} X_1^{j_1} X_2^{j_2}$$

with $a_{j_1,j_2} \in \mathbf{Z}$. For some fixed i_1, i_2, the equation

$$P_{i_1,i_2}(\alpha, \alpha) = 0$$

is a linear equation in the variables a_{j_1,j_k} with coefficients in $\mathbf{Q}(\alpha)$. If we choose a basis in $\mathbf{Q}(\alpha)$ over $\mathbf{Q}$, this becomes a system of $d = [\mathbf{Q}(\alpha) : \mathbf{Q}]$ linear equations over $\mathbf{Q}$.

If we want to find a P such that

$$I_P(\alpha, \alpha) \geq I$$

for some $I \in \mathbf{R}_{\geq 0}$, then we need to solve

$$d \cdot \left| \Big\{ (i_1, i_2) : i_1 \log q_1 + i_2 \log q_2 < I \Big\} \right| = d(1 + o(1)) \frac{I^2}{2 \log(q_1) \log(q_2)}$$

many equations. To obtain the last estimate, note that the pair (i_1, i_2) satisfies $i_1 \log(q_1) + i_2 \log(q_2) < I$ if and only if it belongs to the triangle whose vertices are $(0, 0)$, $(I/\log q_1, 0)$ and $(0, I/\log q_2)$, and $I^2/2 \log(q_1) \log(q_2)$ is the area of this trinagle.

We have $(n_1 + 1)(n_2 + 1)$ many variables, so by linear algebra, we can find a non-zero solution P provided

$$(n_1 + 1)(n_2 + 1) > d \cdot \left| \Big\{ (i_1, i_2) : i_1 \log q_1 + i_2 \log q_2 < I \Big\} \right|.$$

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

We are free to choose our parameters n_1, n_2 and I. To get the desired contradiction between (12) and (14) we need

$$n_1 \log q_1 + n_2 \log q_2 \leq (\sqrt{2d} + \varepsilon) I.$$

So that we can construct P, we need

$$n_1 n_2 \geq (1 + \varepsilon') \frac{d I^2}{2 \log(q_1) \log(q_2)}$$

for some $\varepsilon' > 0$. We can satisfy both constraints by taking

$$n_j \approx \frac{\sqrt{2d} + \varepsilon}{2} \cdot \frac{I}{\log q_j}$$

for $j = 1, 2$.

To prove Theorem 30, we need to carry out the following tasks.

- Construct a polynomial P with large index at (α, α) as described above. 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 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 $P(p_1/q_1, p_2/q_2)$ 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 $P(p_1/q_1, p_2/q_2) = 0$. However, what is possible to show instead is that P cannot have large index at $(p_1/q_1, p_2/q_2)$. This is as good as if $P(p_1/q_1, p_2/q_2)$ did not vanish, because we can exchange P for a suitable derivative. By doing this, we decrease the index at $(\alpha, \dots, \alpha)$, but by not too much, so we can afford it. This should be thought of as a converse for Siegel's Lemma.

  What we show is that if q_1, q_2 are large and P have large index at $(p_1/q_1, p_2/q_2)$ (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,i_2} and the fact that we divide by the factorials $i_1!i_2!$ 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,i_2} also has integer coefficients. In fact, the coefficient of $X_1^{m_1-i_1}X_2^{m_2-i_2}$ in P_{i_1,i_2} is

$$\binom{m_1}{i_1}\binom{m_2}{i_2}$$

times the coefficient of $X_1^{m_1}X_k^{m_2}$ in P. This leads to

$$H(P_{i_1,i_2}) \leq 2^{n_1+n_2} H(P), \tag{15}$$

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 $n_1!n_2!$ in place of $2^{n_1+n_2}$. This is significant, because $2^{n_1+n_2}$ will always be much smaller than $q_j^{n_j}$

provided q_j is sufficiently large and $q_1^{n_1} \approx q_2^{n_2}$. However, we cannot say the same about $n_1!n_2!$.

3.1. **Some historical remarks (Non-examinable).** 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. Interestingly, Siegel obtained a slightly weaker bound than Theorem 30. Instead of the exponent $\sqrt{2d} + \varepsilon$, he obtained

$$\min_{s=1,\ldots,d-1} (s + \frac{d}{s+1} + \varepsilon),$$

which is always a little less than $2\sqrt{d}$. The exponent $\sqrt{2d} + \varepsilon$ was first obtained by Dyson via a different method, which we will comment on in Section 3.3. As we said in the introduction, the optimal exponent $2+\varepsilon$ was achieved by Roth who used polynomials in many variables. The new significant difficulty is estimating the index of P at the rational approximations, which is more difficult to prove in several variables.

Thue's proof was based on the simple estimate

$$\left|\frac{p}{q} - \frac{r}{s}\right| \geq \frac{1}{qs},$$

which holds for any pair of distinct rational numbers. His plan was to construct suitable rational approximations of α 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 α, we can produce a sequence of better and better rational approximations. These converge very fast to α, 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 α 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 other 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

$$|2^{1/3} - p/q| \le \frac{1}{s^{2.5+\varepsilon}}$$

and $q \le s^{1.5+\varepsilon}/2$ indeed, that would imply

$$|2^{1/3} - r/s| \ge \frac{1}{qs} - \frac{1}{s^{2.5+\varepsilon}} \ge \frac{1}{s^{2.5+\varepsilon}}.$$

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 $p/q \ne r/s$, 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

$$\frac{(d^m/dx^m)R_n(p_0/q_0)}{(d^m/dx^m)Q_n(p_0/q_0)} \ne r/s$$

for some m that is not so large.

For a more complete and very nicely written discussion, see [9, 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 a 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 $H(L)$ of L is defined as

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

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

Siegel's Lemma 31. Let K be a number field and let $D = [K : \mathbf{Q}]$. Let $M, N \in \mathbf{Z}_{>0}$ with $MD < N$, and let $\mathcal{H} \in \mathbf{R}_{\ge 1}$. Let $L_1, \ldots, L_M \in K[x_1, \ldots, x_N]$ be linear forms with $H(L_j) \le \mathcal{H}$ for all j.

Then there are $x_1, \ldots, x_N \in \mathbf{Z}$ not all 0 with

$$L_j(x_1, \ldots, x_N) = 0, \qquad j = 1, \ldots, M$$

and

$$|x_j| \le (N\mathcal{H})^{\frac{DM}{N-DM}}.$$

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 $\mathcal{H} \geq 1$ cannot be replaced by $\mathcal{H} \geq 0$.
- Note that DM is the number of linear equations over $\mathbf{Q}$ 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 $|x_j| \leq N\mathcal{H}$.
- 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 [2, 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 32. *Let $\alpha \in \overline{\mathbf{Q}}$. Let $n_1, n_2 \in \mathbf{Z}_{\geq 1}$, let $w_1, w_2 \in \mathbf{R}_{>0}$ and let $I, \delta \in \mathbf{R}_{>0}$ be such that*

$$|\{(i_1, i_2) \in \mathbf{Z}^k_{\geq 0} : i_1 w_1 + i_2 w_2 \leq I\}| \leq \frac{(n_1+1)(n_2+1)}{(1+\delta)[\mathbf{Q}(\alpha) : \mathbf{Q}]}.$$

Then there is $P \neq 0 \in \mathbf{Z}[X_1, X_2]$ of degree at most n_j in X_j such that

$$H(P) \leq (4H(\alpha))^{\delta^{-1}(n_1+n_2)}$$

and

$$I_P(\alpha, \alpha; w_1, w_2) \geq I.$$

Proof. We search for P in the form

$$P = \sum_{j_1=0}^{n_1} \sum_{j_2=0}^{n_2} a_{j_1,j_2} X_1^{j_1} X_2^{j_2}.$$

For $(i_1, i_2) \in \mathbf{Z}^2_{\geq 0}$, we consider the linear form

$$L_{i_1,i_2}(a_{j_1,j_2}) = \sum_{j_1,j_2} \binom{j_1}{i_1}\binom{j_2}{i_2} \alpha^{j_1+j_2-i_1-i_2} a_{j_1,j_2}.$$

Here the range of j_l is from 0 to n_l for $l = 1, 2$. We note that $\binom{j_l}{i_l} = 0$ if $i_l > j_l$. The point of this is, of course, that $P_{i_1,i_k}(\alpha, \alpha) = 0$ is equivalent to

$$L_{i_1,i_2}(a_{j_1,j_2}) = 0.$$

We can estimate $|L_{i_1,i_2}|_v$ similarly to the proof of Proposition 28. For finite places v, we write

$$|L_{i_1,i_2}|_v \leq \max(1, |\alpha|_v)^{n_1+n_2}.$$

For the infinite places v, we write

$$|L_{i_1,i_2}|_v \le 2^{n_1+n_2} \max(1, |\alpha|_v)^{n_1+n_2}.$$

This is based on $\binom{j_l}{i_l} \le 2^{n_l}$. Multiplying together these bounds, we get

$$H(L_{i_1,i_2}) \le (2H(\alpha))^{n_1+n_2}.$$

We can apply Siegel's Lemma with $D = [\mathbf{Q}(\alpha) : \mathbf{Q}]$, $\mathcal{H} = (2H(\alpha))^{n_1+n_2}$ and

$$N = (1 + n_1)(1 + n_2) \le 2^{n_1+n_2}.$$

By the assumption of the corollary, the number M of linear forms we need to consider satisfies $(1 + \delta)DM \le N$. Then

$$\frac{DM}{N - DM} \le \frac{DM}{(1 + \delta)DM - DM} = \delta^{-1}$$

and hence we obtain a non-zero solution with

$$|a_{j_1,j_2}| \le (N\mathcal{H})^{\delta^{-1}} \le (4H(\alpha))^{\delta^{-1}(n_1+n_2)}.$$

This completes the proof. □

Now we turn to the proof of Siegel's Lemma. We will give the details only in the special case $K = \mathbf{Q}$. The general case has 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 $K = \mathbf{Q}$. 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 $\mathcal{H}$.

Let

$$Y = \left\lfloor (N\mathcal{H})^{\frac{M}{N-M}} \right\rfloor,$$

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 that the coefficients of L_j are at most $\mathcal{H}$, we get

$$\max_{y_1,\ldots,y_N} L_j(y_1, \ldots, y_N) - \min_{y_1,\ldots,y_N} L_j(y_1, \ldots, y_N) \le N\mathcal{H}Y.$$

There are $(Y + 1)^N$ choices for $y_1, \ldots, y_N$, and $L_j(y_1, \ldots, y_N)$ may take at most $N\mathcal{H}Y + 1 \le N\mathcal{H}(Y + 1)$ different values for each j. (Here we used $\mathcal{H} \ge 1$.) Note that

$$(Y + 1)^{N-M} > \left((N\mathcal{H})^{\frac{M}{N-M}}\right)^{N-M},$$

hence $(Y + 1)^N > (N\mathcal{H}(Y + 1))^M$. By the box principle, there are two distinct vectors

$$(y_1, \ldots, y_N) \ne (z_1, \ldots, z_N) \in \{0, \ldots, Y\}^N$$

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

$$|x_i| \leq Y \leq (N\mathcal{H})^{\frac{M}{N-M}}.$$

□

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 [9, Proposition 14.12]. We consider the map

$$\Phi : \alpha \mapsto (v(\alpha))_{v \in M_{K,\infty}},$$

which maps K to a product of copies of $\mathbf{R}$ and $\mathbf{C}$. The dimension of this product over $\mathbf{R}$ is D, so we identify it with $\mathbf{R}^D$.

Similarly to the proof in the $K = \mathbf{Q}$ case, we consider the vectors

$$A := \{(L_j(y_1, \ldots, y_N))_{j=1,\ldots,M} : y_i = 0, \ldots, Y\} \subset K^M,$$

where Y is a fixed parameter. We can also show that the coordinate(s) of the points in $\Phi(A)$ corresponding to the embedding v fall in an interval of length $N|L_j|_v Y$ for each $v \in M_{K,\infty}$. This allows us to confine $\Phi(A)$ in a certain box in $\mathbf{R}^{DM}$.

Now we would like to have an upper bound on the number of possible vectors in A whose Φ image falls in that box. To this end, we will use the product formula

$$\prod_{v \in M_{K,\infty}} |\alpha|_v^{d_v} = \prod_{v \in M_{K,f}} |\alpha|_v^{-d_v}$$

for

$$\alpha = L_j(y_1, \ldots, y_N) - L_j(z_1, \ldots, z_N)$$

for some fixed j. For any α of this form, we can write

$$|\alpha|_v \leq |L_j|_v$$

for all finite place v, so we get

$$\prod_{v \in M_{K,\infty}} |\alpha|_v^{d_v} \geq \prod_{v \in M_{K,f}} |L_j|_v^{-d_v}. \tag{16}$$

This allows us to place a suitable small box around each point in $\Phi(A)$, which will be pairwise disjoint. In fact, we can choose these boxes in many ways, (16) only requires that the product of the side lengths of the box belonging to the j'th $\mathbf{R}^D$ in $\mathbf{R}^{DM}$ does not exceed $\prod_{v \in M_{K,f}} |L_j|_v^{-1}$.

Now we can estimate $|\Phi(A)|$ by comparing the volume of the big box containing all these points against the volume of the small boxes around each point. 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 $K = \mathbf{Q}$.

3.3. **Non-vanishing.** The purpose of this section is to show that a polynomial $P \in \mathbf{Z}[x_1, x_2]$ with not too large coefficients and satisfying some further hypothesis cannot have large index at a rational point $(p_1/q_1, p_2/q_2)$. The precise statement is the following.

Proposition 33. *For all $\varepsilon > 0$, there is a constant $C = C(\varepsilon)$ such that the following holds. Let $n_1, n_2 \in \mathbf{Z}_{\geq 1}$. Let $p_1/q_1, p_2/q_2 \in \mathbf{Q}$ such that $\log q_2 \geq C \log q_1$ and*

$$\exp(n_1 + n_2) \leq q_j^{n_j/C}.$$

for $j = 1, 2$. Let $P \neq 0 \in \mathbf{Z}[X_1, X_2]$ be a polynomial of degree at most n_j in X_j for $j = 1, 2$. Suppose

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

for $j = 1, 2$. Then

$$I_P(p_1/q_1, p_2/q_2; \log q_1, \log q_2) \leq \varepsilon(n_1 \log q_1 + n_2 \log q_2).$$

Some remarks are in order.

- When we write down "let $p/q \in \mathbf{Q}$" 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 $q_1^{n_1}$ and $q_2^{n_2}$ are roughly of the same size and q_1 is very large. Then the condition on $H(P)$ is satisfied always when $H(P) \leq \widetilde{C}^{n_1 + \dots + n_k}$, where $\widetilde{C}$ 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. Therefore this lemma should be thought of as a converse to Siegel's lemma.
- This 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 (α, α) and at $(p_1/q_1, p_2/q_2)$. This statement does not require any assumption on $H(P)$ or the size of q_j. In fact, it can be formulated entirely in the setting of polynomials with complex coefficients. This has been done for polynomials in two variables by Dyson, and 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 [10].

- 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 $R, Q \in \mathbf{Z}[X_1]$, 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 $p_1/q_1 \in \mathbf{Q}$. If R and Q has low degree compared to ε, 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 first explain the strategy of proof. We begin with a simple auxiliary result that will help us to make computations with the index of a polynomial.

Lemma 34. *Let $F, F^{(1)}, F^{(2)} \in \mathbf{Z}[X_1, X_2]$, let $i_1, i_2 \in \mathbf{Z}_{\geq 0}$, let $\alpha_1, \alpha_2 \in \mathbf{R}$, and let $w_1, w_2 \in \mathbf{R}_{>0}$. Then the following holds*

$$\begin{aligned} I_{F_{i_1,i_2}}(\alpha_1, \alpha_2) \geq & I_F(\alpha_1, \alpha_2) - i_1 w_1 - i_2 w_2, \\ I_{F^{(1)}+F^{(2)}}(\alpha_1, \alpha_2) \geq & \min_{j=1,2} I_{F^{(j)}}(\alpha_1, \alpha_2), \\ I_{F^{(1)}F^{(2)}}(\alpha_1, \alpha_2) = & I_{F^{(1)}}(\alpha_1, \alpha_2) + I_{F^{(2)}}(\alpha_1, \alpha_2). \end{aligned}$$

In these formulas the index is always understood to be with respect to w_1, w_2.

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 (α_1, α_2). We omit the details.

Proposition 33 is easy to prove when P is of the special form

$$P(X_1, X_2) = F(X_1)G(X_2) \tag{17}$$

for some polynomials F and G. Indeed, we could estimate the heights of F and G in terms of $H(P)$ and use this to bound the indices of F and G. This can be done, for example, by using that a polynomial with integer coefficients vanishing at a rational point p/q to order m must have leading coefficient at least q^m. Then the third item of Lemma 34 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

$$P = F^{(1)}G^{(1)} + \ldots + F^{(h)}G^{(h)} \tag{18}$$

for some $h \in \mathbf{Z}_{\geq 1}$, $F^{(j)} \in \mathbf{Z}[X_1]$ and $G^{(j)} \in \mathbf{Z}[X_2]$. Indeed, we could just write P as a polynomial in X_2 with coefficients in $\mathbf{Z}[X_1]$, and we get an identity of the above form.

Now the idea is to use the identity (18) to replace P by another polynomial that admits a factorization of the form (17). 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

$$F^{(1)}\frac{\partial}{\partial x_2}G^{(1)} + F^{(2)}\frac{\partial}{\partial x_2}G^{(2)}$$

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 $(\partial/\partial x_2)G^{(2)}$ and subtract from it $G^{(2)}$ times the first one. We obtain

$$F^{(1)}(G^{(1)}\frac{\partial}{\partial x_2}G^{(2)} - \frac{\partial}{\partial x_2}G^{(1)} \cdot G^{(2)}).$$

By Lemma 34, 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

$$\begin{vmatrix} P & G^{(2)} & \dots & G^{(h)} \\ \frac{\partial}{\partial X_2}P & \frac{\partial}{\partial X_2}G^{(2)} & \dots & \frac{\partial}{\partial X_2}G^{(h)} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^{h-1}}{\partial X_2^{h-1}}P & \frac{\partial^{h-1}}{\partial X_2^{h-1}}G^{(2)} & \dots & \frac{\partial^{h-1}}{\partial X_2^{h-1}}G^{(h)} \end{vmatrix}. \tag{19}$$

Using Lemma 34, this has index as large as $\frac{\partial^{h-1}}{\partial X_2^{h-1}}P$, and using (18) for the first column, we get that (19) has the factorization

$$F^{(1)} \cdot \begin{vmatrix} G^{(1)} & G^{(2)} & \dots & G^{(h)} \\ \frac{\partial}{\partial X_2}G^{(1)} & \frac{\partial}{\partial X_2}G^{(2)} & \dots & \frac{\partial}{\partial X_2}G^{(h)} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^{h-1}}{\partial X_2^{h-1}}G^{(1)} & \frac{\partial^{h-1}}{\partial X_2^{h-1}}G^{(2)} & \dots & \frac{\partial^{h-1}}{\partial X_2^{h-1}}G^{(h)} \end{vmatrix}.$$

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.

In the proof of Proposition 33, we will consider the determinant

$$\begin{vmatrix} P_{0,0} & P_{0,1} & \dots & P_{0,h-1} \\ P_{1,0} & P_{1,1} & \dots & P_{1,h-1} \\ \vdots & \vdots & & \vdots \\ P_{h-1,0} & P_{h-1,1} & \dots & P_{h-1,h-1} \end{vmatrix}.$$

Recall that

$$P_{i,j} = \frac{1}{i!j!}\frac{\partial^{i+j}P}{\partial x_1^i \partial x_2^j}.$$

This has a factorization of the form

$$\begin{vmatrix} F^{(1)} & F^{(2)} & \dots & F^{(h)} \\ F_1^{(1)} & F_1^{(2)} & \dots & F_1^{(h)} \\ \vdots & \vdots & & \vdots \\ F_{h-1}^{(1)} & F_{h-1}^{(2)} & \dots & F_{h-1}^{(h)} \end{vmatrix} \cdot \begin{vmatrix} G^{(1)} & G_1^{(1)} & \dots & G_{h-1}^{(1)} \\ G^{(2)} & G_1^{(2)} & \dots & G_{h-1}^{(2)} \\ \vdots & \vdots & & \vdots \\ G^{(h)} & G_1^{(h)} & \dots & G_{h-1}^{(h)} \end{vmatrix}.$$

The determinants appearing in the factorization are called Wronskians, which play an important role in the theory of linear ODE's. We will need to show that these determinants does not vanish identically, otherwise the factorisation is useless. To this end, we will need at least that the functions $F_1, \dots, F_h$ and $G_1, \dots, G_h$ are linearly independent over $\mathbf{Q}$. According to the next lemma, this is also enough.

Lemma 35. *Let $F^{(1)}, \dots, F^{(h)} \in \mathbf{Q}[X_1]$ be $\mathbf{Q}$-linearly independent polynomials. Then*

$$\begin{vmatrix} F_0^{(1)} & F_0^{(2)} & \dots & F_0^{(h)} \\ F_1^{(1)} & F_1^{(2)} & \dots & F_1^{(h)} \\ \vdots & \vdots & & \vdots \\ F_{h-1}^{(1)} & F_{h-1}^{(2)} & \dots & F_{h-1}^{(h)} \end{vmatrix} \neq 0.$$

Proof. By the properties of determinants and differentiation, non-vanishing of the Wronskian does not change if we replace $F^{(j)}$ by $aF^{(j)} - bF^{(i)}$ for some $i \neq j$ and $a, b \in \mathbf{Q}$ 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

$$\begin{vmatrix} X_1^{m_1} & X_1^{m_2} & \dots & X_1^{m_h} \\ \frac{\partial}{\partial X_1} X_1^{m_1} & \frac{\partial}{\partial X_1} X_1^{m_2} & \dots & \frac{\partial}{\partial X_1} X_1^{m_h} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^{h-1}}{(h-1)!\partial X_1^{h-1}} X_1^{m_1} & \frac{\partial^{h-1}}{(h-1)!\partial X_1^{h-1}} X_1^{m_2} & \dots & \frac{\partial^{h-1}}{(h-1)!\partial X_1^{h-1}} X_1^{m_h} \end{vmatrix} \tag{20}$$

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 (20) $\neq 0$. To this end, we write

$$(20) = \begin{vmatrix} \binom{m_1}{0} & \binom{m_2}{0} & \dots & \binom{m_h}{0} \\ \binom{m_1}{1} & \binom{m_2}{1} & \dots & \binom{m_h}{1} \\ \vdots & \vdots & & \vdots \\ \binom{m_1}{h-1} & \binom{m_2}{h-1} & \dots & \binom{m_h}{h-1} \end{vmatrix} \cdot X_1^M,$$

where $M \in \mathbf{Z}$ is a number whose value is immaterial. The binomial coefficient $\binom{m}{j}$ can be considered a polynomial of degree j in the variable m. These are linearly independent polynomials of degree at most $h-1$ for $j = 0, \dots, h-1$, so no non-trivial linear combination of them 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. $\square$

Proof of Proposition 33. We assume to the contrary that P and $p_1/q_1, p_2/q_2$ 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 ε.

We write P in the form

$$P = F^{(1)}(X_1)G^{(1)}(X_2) + \cdots + F^{(h)}(X_1)G^{(h)}(X_2) \tag{21}$$

for some $1 \le h \le n_k$. We assume, as we may, that $F^{(1)}, \ldots, F^{(h)}$ and $G^{(1)}, \ldots, G^{(h)}$ are $\mathbf{Q}$-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 (21) with one less term. This process clearly has to terminate at some point.

We consider the matrix

$$\mathcal{P}(x_1, x_2) = \begin{vmatrix} P_{0,0} & P_{0,1} & \ldots & P_{0,h-1} \\ P_{1,0} & P_{1,1} & \ldots & P_{1,h-1} \\ \vdots & \vdots & & \vdots \\ P_{h-1,0} & P_{h-1,1} & \ldots & P_{h-1,h-1} \end{vmatrix},$$

which has a factorization of the form

$$\mathcal{F}(x_1) \cdot \mathcal{G}(x_2) = \begin{vmatrix} F^{(1)} & F^{(2)} & \ldots & F^{(h)} \\ F_1^{(1)} & F_1^{(2)} & \ldots & F_1^{(h)} \\ \vdots & \vdots & & \vdots \\ F_{h-1}^{(1)} & F_{h-1}^{(2)} & \ldots & F_{h-1}^{(h)} \end{vmatrix} \cdot \begin{vmatrix} G^{(1)} & G_1^{(1)} & \ldots & G_{h-1}^{(1)} \\ G^{(2)} & G_1^{(2)} & \ldots & G_{h-1}^{(2)} \\ \vdots & \vdots & & \vdots \\ G^{(h)} & G_2^{(h)} & \ldots & G_{h-1}^{(h)} \end{vmatrix},$$

By Lemma 35, $\mathcal{F}, \mathcal{G} \neq 0$, hence also $\mathcal{P} \neq 0$.

In what follows, we estimate the degrees and heights of $\mathcal{F}$ and $\mathcal{G}$ in terms of the degrees and height of P. We use this to bound their indices at p_1/q_1 and p_2/q_2 respectively. This in turn will yield a bound on the index of $\mathcal{P}$ at $(p_1/q_1, p_2/q_2)$, 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 $j = 1$ and in $\mathcal{G}$ for $j = 2$.

The height of each entry of $\mathcal{P}$, is at most $2^{n_1+n_2}H(P)$. 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 $(n_1+1)^h(n_2+1)^h$ ways of choosing these monomials. We have therefore

$$H(\mathcal{P}) \le h! \cdot (n_1+1)^h(n_2+1)^h \cdot (2^{n_1+n_2}H(P))^h \le 2^{3hn_1+3hn_2} \cdot H(P)^h.$$

Here we used $h \le \min(n_1, n_2)$. Since $\mathcal{F}$ and $\mathcal{G}$ has disjoint sets of variables, we have $H(\mathcal{P}) = H(\mathcal{F})H(\mathcal{G})$, and hence

$$H(\mathcal{F}), H(\mathcal{G}) \le 2^{3hn_1+3hn_2} \cdot H(P)^h.$$

Since P satisfies the hypotheses of the proposition, we have

$$H(\mathcal{F}), H(\mathcal{G}) \le q_j^{\log(8)hn_j/C} \cdot q_j^{hn_j/C}$$

for $j = 1, 2$.

We estimate the index of $\mathcal{F}$ at p_1/q_1. Suppose p_1/q_1 is a zero of order m of $\mathcal{F}$. By Gauss's lemma, $\mathcal{F}$ is divisible by $(q_1X_1 - p_1)^m$. Then the leading coefficient of $\mathcal{F}$ is divisible by q_1^m, hence $H(\mathcal{F}) \ge q_1^m$. By this, and a similar argument for $\mathcal{G}$, we get

$$I_{\mathcal{F}}(p_1/q_1; \log q_1) \le \log H(\mathcal{F}) \le hn_1 \log q_1/C$$
$$I_{\mathcal{G}}(p_2/q_2; \log q_2) \le \log H(\mathcal{G}) \le hn_2 \log q_2/C.$$

By Lemma 34, we conclude

$$I_{\mathcal{P}}(p_1/q_1, p_2/q_2; \log q_1, \log q_2) \le h(n_1 \log q_1 + n_2 \log q_2)/C.$$

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 34, the index of each entry in the first l columns of $\mathcal{P}$ is at least

$$\varepsilon(n_1 \log q_1 + n_2 \log q_2) - (h-1)\log q_1 - (l-1)\log q_2.$$

If C is sufficiently large in terms of ε, then $\log q_1 < \varepsilon/10 \cdot \log q_2$ and if $l \le \varepsilon/10 \cdot h + 1$, then the above index is at least

$$(\varepsilon/2)(n_1 \log q_1 + n_2 \log q_2).$$

There are at least $(\varepsilon/10)h$ columns of $\mathcal{P}$, where this bound holds. Now we expand the determinant $\mathcal{P}$ and use Lemma 34 again to conclude

$$I_{\mathcal{P}}(p_1/q_1, p_2/q_2; \log q_1, \log q_2) \ge (\varepsilon^2/20)h(n_1 \log q_1 + n_2 \log q_2).$$

This clearly contradicts our upper bound, provided C is sufficiently large in terms of ε. $\square$

3.4. **Completing the proof.** We complete the proof of Theorem 30. Let α and ε be as in the theorem, and let $\varepsilon_0 > 0$ be a sufficiently small number, which we will choose later depending on α and ε. Let $d = [\mathbf{Q}(\alpha) : \mathbf{Q}]$. Suppose to the contrary that there are infinitely many rationals $p/q \in \mathbf{Q}$ such that $|\alpha - p/q| < q^{-\sqrt{2d}-\varepsilon}$. Fix two among these, p_1/q_1 and p_2/q_2 such that

$$\log q_2 \ge C \log q_1 \tag{22}$$

$$\log q_1 \ge C\varepsilon_0^{-1}, \tag{23}$$

where C is the constant in Proposition 33 applied with ε_0 in place of ε.

Choose two integers n_1, n_2, which are sufficiently large in terms of α, ε, ε_0, q_1 and q_2, and such that

$$|n_1 \log q_1 - n_2 \log q_2| \leq \log q_1.$$

We set

$$I = \frac{n_1 \log q_1 + n_2 \log q_2}{\sqrt{2d} + \varepsilon/10}.$$

If $i_1, i_2 \in \mathbf{Z}_{\geq 0}$ are such that $i_1 \log q_1 + i_2 \log q_2 \leq I$, then the entire unit square whose lower left corner is (i_1, i_2) is contained in the triangle whose vertices are

$$(0,0), \quad ((I + \log q_1 + \log q_2)/\log q_1, 0), \quad (0, (I + \log q_1 + \log q_2)/\log q_2).$$

Therefore,

$$\begin{aligned} |\{(i_1, i_2) \in \mathbf{Z}^2_{\geq 0} : i_1 \log q_1 + i_2 \log q_2 \leq I\}| &\leq \frac{(I + \log q_1 + \log q_2)^2}{2 \log q_1 \log q_2} \\ &\leq \frac{(n_1+1)(n_2+1)}{(1+\delta)d} \end{aligned}$$

for a suitable $\delta > 0$ depending on ε provided n_1 and n_2 are sufficiently large.

Therefore, we may use Corollary 32, the corollary to Siegel's lemma, and conclude that there is some $P \neq 0 \in \mathbf{Z}[X_1, X_2]$ of degree at most n_j in x_j such that $H(P) \leq (4H(\alpha))^{\delta^{-1}(n_1+n_2)}$ and $I_P(\alpha, \alpha; \log q_1, \log q_2) \geq I$.

Now we use Proposition 33 to bound the index of P at $(p_1/q_1, p_2/q_2)$. If ε_0 is chosen sufficiently small in terms of δ, then the condition $\log q_1 \geq C\varepsilon_0^{-1}$ and the choice of n_1, n_2 implies

$$\exp(n_1 + n_2), H(P) \leq q_j^{n_j/C}$$

for $j = 1, 2$. Then the proposition applies and

$$I_P(p_1/q_1, p_2/q_2; \log q_1, \log q_2) \leq \varepsilon_0(n_1 \log q_1 + n_2 \log q_2).$$

It follows that there is a suitable partial derivative $\widetilde{P}$ of P that satisfies

$$H(\widetilde{P}) \leq 2^{n_1+n_2} H(P) \leq (8H(\alpha))^{\delta^{-1}(n_1+n_2)},$$

$$I_{\widetilde{P}}(\alpha, \alpha; \log q_1, \log q_2) \geq I - \varepsilon_0(n_1 \log q_1 + n_2 \log q_2) \geq \frac{n_1 \log q_1 + n_2 \log q_2}{\sqrt{2d} + \varepsilon/5},$$

$$\widetilde{P}(p_1/q_1, p_2/q_2) \neq 0$$

provided ε_0 is sufficiently small in terms of ε.

The last property implies the lower bound

$$|\widetilde{P}(p_1/q_1, p_2/q_2)| \geq \frac{1}{q_1^{n_1} q_2^{n_2}}.$$

We use Taylor expansion around (α, α) to produce an upper bound. We write

$$\widetilde{P}(p_1/q_1, p_2/q_2) = \sum_{i_1,i_2} \widetilde{P}_{i_1,i_2}(\alpha, \alpha)(\alpha - p_1/q_1)^{i_1}(\alpha - p_2/q_2)^{i_2}.$$

The number of terms in the expansion is less than $(n_1 + 1)(n_2 + 1) \leq 2^{n_1+n_2}$. The coefficients of $\widetilde{P}_{i_1,i_2}$ are bounded by $(16H(\alpha))^{\delta^{-1}(n_1+n_2)}$, and we have $|\alpha| < H(\alpha)^d$. Thus

$$\begin{aligned}|\widetilde{P}_{i_1,i_2}(\alpha, \alpha)| &\leq (n_1 + 1)(n_2 + 1)(16H(\alpha))^{\delta^{-1}(n_1+n_2)} H(\alpha)^{d(n_1+n_2)}\\ &\leq (32H(\alpha))^{\delta^{-1}(d+1)(n_1+n_2)}.\end{aligned}$$

Finally, for each (i_1, i_2) with $P_{i_1,i_2}(\alpha, \alpha) \neq 0$, we have

$$\begin{aligned}|\alpha - p_1/q_1|^{i_1}|\alpha - p_2/q_2|^{i_2} &\leq q_1^{-(\sqrt{2d}+\varepsilon)i_1} q_2^{-(\sqrt{2d}+\varepsilon)i_2}\\ &\leq \exp(-(\sqrt{2d} + \varepsilon)I_{\widetilde{P}}(\alpha, \alpha; \log q_1, \log q_2))\\ &\leq \Big(\frac{1}{q_1^{n_1} q_2^{n_2}}\Big)^{\frac{\sqrt{2d}+\varepsilon}{\sqrt{2d}+\varepsilon/5}}\end{aligned}$$

Putting all these together, we obtain the upper bound

$$\begin{aligned}|\widetilde{P}(p_1/q_1, p_2/q_2)| &\leq 2^{n_1+n_2}(32H(\alpha))^{\delta^{-1}(d+1)(n_1+n_2)}\Big(\frac{1}{q_1^{n_1} q_2^{n_2}}\Big)^{\frac{\sqrt{2d}+\varepsilon}{\sqrt{2d}+\varepsilon/5}}\\ &< \frac{1}{q_1^{n_1} q_2^{n_2}}\end{aligned}$$

provided q_1 is sufficiently large in terms of α, δ and ε, which we can force by taking ε_0 to be small enough. This contradicts our previous lower bound, which completes the proof.

3.5. **Towards Roth's theorem (Non-examinable).** In this section, we discuss how the above argument can be improved to obtain the exponent $2 + \varepsilon$, which was achieved by Roth.

For this, one needs to work with auxiliary polynomials in k variables instead of just 2, where k is chosen depending on ε and d. Most of the argument carries over without much trouble. The main exception is the non-vanishing result the k variable version of which is known as Roth's lemma. The proof of this is the key difficulty which prevented the an improvement of the exponent $\sqrt{2d}+\varepsilon$ before Roth's work.

Roth's Lemma 36. For all $\varepsilon > 0$ and $k \in \mathbf{Z}_{\geq 1}$ there is a constant $C = C(\varepsilon, k)$ such that the following holds. Let $n_1, \ldots, n_k \in \mathbf{Z}_{\geq 1}$. Let $p_1/q_1, \ldots, p_k/q_k \in \mathbf{Q}$ 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 $P \neq 0 \in \mathbf{Z}[X_1, \ldots, X_k]$ 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

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

The proof of Roth's lemma is by induction on the number of variables k. We have seen the proof for $k = 1$ and 2. The induction step is based on similar ideas to the $k = 2$ case. The main difference is that in the factorization $\mathcal{P} = \mathcal{F}\mathcal{G}$, one of the factors is a polynomial in more than one variable, and it is a "generalized Wronskian", meaning that we use a suitable sequence of partial derivatives in place of $1, d/dX, \ldots, d^{h-1}/dX^{h-1}$.

Next we discuss how having more variables helps with improving the exponent. Suppose that $p_1/q_1, \ldots, p_k/q_k \in \mathbf{Q}$ are such that

$$|\alpha - p_j/q_j| < q_j^{-t}$$

for some exponent $t > 0$. Using Siegel's lemma, we construct an auxiliary polynomial $P(x_1, \ldots, x_k)$ of degree n_j in x_j with not too big coefficients such that

$$I_P(\alpha, \ldots, \alpha; \log q_1, \ldots, \log q_k) \geq I$$

for an appropriately chosen I. Here we choose the degrees in such a way that $n_j \log q_j$ are roughly the same for all j.

Using Roth's lemma, we can pass to a partial derivative, which we continue to denote by P, such that $P(p_1/q_1, \ldots, p_k/q_k) \neq 0$. The lower bound

$$|P(p_1/q_1, \ldots, p_k/q_k)| > \frac{1}{q_1^{n_1} \cdots q_k^{n_k}}$$

follows. Using a Taylor expansion around $(\alpha, \ldots, \alpha)$ we get an upper for the same roughly equal to $\exp(-tI)$. We get a contradiction if

$$tI > n_1 \log q_1 + \ldots + n_k \log q_k.$$

Now the question is how large we may take I so that

$$|\{(i_1, \ldots, i_k) : i_1 \log q_1 + \ldots + i_k \log q_k < I\}| \leq \frac{(n_1 + 1) \cdots (n_k + 1)}{2d},$$

say. This has a probabilistic interpretation. Consider i_j for $j = 1, \ldots, k$ to be independent uniform random variables in $0, \ldots, n_j$. By the law of large numbers,

$$\frac{i_1 \log q_1 + \ldots + i_k \log q_k}{n_1 \log q_1 + \ldots + n_k \log q_k}$$

will be close to 1/2 with arbitrarily high probability provided we choose k sufficiently large. (Here we used that the $n_j \log q_j$ are roughly equal.) This shows that only a small proportion of the indices will satisfy

$$i_1 \log q_1 + \ldots + i_k \log q_k < I$$

provided we take $I = (n_1 \log q_1 + \ldots + n_k \log q_k)/t$ with any $t > 2$.

For a more detailed argument we refer to Cassels's book [5, Chapter VI].

4. Gelfond-Schneider – proof

In this chapter, we prove the Gelfond-Schneider theorem about the linear independence of two logarithms of algebraic numbers. We give full details in the case when the logarithms are real. At the end of the section, we give brief remarks about extensions of the method. We follow Schneider's method with a modification introduced by Laurent, which uses large determinants instead of auxiliary functions. We borrow heavily from Waldschmidt [13, Chapter 2].

Our approach in the previous section can be summarized as follows. We considered functions $\varphi_1, \ldots, \varphi_L$, which was an enumeration of the monomials $x_1^{j_1} x_2^{j_2}$ for some pairs of indices (j_1, j_2). Then we used Siegel's lemma to find coefficients $a_1, \ldots, a_L$ such that the auxiliary function

$$P = a_1\varphi_1 + \ldots + a_L\varphi_L$$

vanishes to high order at some point u_1, which was $u_1 = (\alpha, \alpha)$ in our case. We were not talking about this, (because we did not need to), but, of course, P also vanishes at all the Galois conjugates of (α, α), which we denote by $u_2, \ldots, u_d$, where $d = [\mathbf{Q}(\alpha) : \mathbf{Q}]$.

Then we considered another point u_{d+1}, which was $(p_1/q_1, p_2/q_2)$, and using an indirect hypothesis and an analytic argument, we proved that $P(u_{d+1})$ is very small. We contrasted this by an arithmetic lower bound coming from Liouville's inequality to show that $P(u_{d+1}) = 0$. Iterating this argument for partial derivatives, shows that P vanishes at u_{d+1} to a certain multiplicity. This yielded a contradiction, because we showed that the polynomial we constructed can vanish at u_{d+1} to very small multiplicity only.

By vanishing at $u_1, \ldots, u_{d+1}$ to certain multiplicities, we mean that for $i = 1, \ldots, L$, there are indices $k(i) \in \{1, \ldots, d+1\}$ and partial differential operators ∂_i such that

$$\partial_i P(u_{k(i)}) = 0 \qquad \text{for } i = 1, \ldots, L. \tag{24}$$

(In the actual proof, the total number monomials and the total number of partial derivatives are not the same number L, but we will ignore this point, at least for now.) So the proof involved showing both the existence of an auxiliary function P satisfying (24) and that an auxiliary function with so much vanishing cannot exist.

We can express (24) in an alternative way that does not involve an auxiliary function. Consider the matrix

$$M = \begin{pmatrix} \partial_1\varphi_1(u_{k(1)}) & \dots & \partial_L\varphi_1(u_{k(L)}) \\ \vdots & \ddots & \vdots \\ \partial_1\varphi_L(u_{k(1)}) & \dots & \partial_L\varphi_L(u_{k(L)}) \end{pmatrix},$$

and observe that we may write (24) as $(a_1, \dots, a_L)M = (0, \dots, 0)$. We see that the existence of an auxiliary function P satisfying (24) is equivalent to $\det M = 0$. A determinant of this form is called an interpolation determinant. When there are no derivatives involved, it is called an alternant.

So the proof can described as proving that $\det M$ is both 0 and nonzero. There is an alternative approach for proving $\det M = 0$. For example, the assumption that (α, α) is close to $(p_1/q_1, p_2/q_2)$ implies that some of the columns of the matrix are very close to each other, and this may yield an upper bound on $\det M$. Since $\det M$ is an algebraic number, we may contrast our upper bound for $\det M$ with the lower bound from Liouville's inequality resulting $\det M = 0$.

This approach runs into some difficulties. Perhaps the most serious one is that our non-vanishing result relied on the coefficients $a_1, \dots, a_L$ being not too large rational integers, and this does not imply $\det M \neq 0$. However, this does not concern us, because we are not aiming for a new proof of Theorem 30, instead we want to prove the Gelfond-Schneider theorem. The purpose of the above discussion was just to motivate what comes next.

Many proofs in this subject can be explained either by constructing an auxiliary function using Siegel's lemma, or by writing down a suitable determinant. There is no fundamental difference between the two approaches. Schneider's original proof involved an auxiliary function, but we will present it using a determinant, an approach introduced by Laurent.

In what follows, we consider two real numbers $\lambda_1, \lambda_2 \neq 0$ such that $\alpha_1 = e^{\lambda_1}$ and $\alpha_2 = e^{\lambda_2}$ are algebraic numbers. We assume to the contrary that $\beta = \lambda_2/\lambda_1$ is algebraic but not rational. This is the negation of a special case of the Gelfond-Schneider theorem, where the logarithms are assumed real.

We fix some positive integer parameters T_0, T_1 and S such that

$$(T_0 + 1)(2T_1 + 1) = (2S + 1)^2 =: L.$$

We consider the "monomials" $x^\tau \exp(t\lambda_1 x)$ for $\tau = 0, \dots, T_0$ and $t = -T_1, \dots, T_1$. We will evaluate them at the points $x = s_1 + \beta s_2$ for $s_1 = -S, \dots, S$ and $s_2 = -S, \dots, S$. Since β is not rational, these are

$(2S+1)^2$ different points. This leads us to consider the alternant
(25)
$$\Delta := \det[(s_1+\beta s_2)^\tau \exp(t\lambda_1(s_1+\beta s_2))]_{\substack{t,\tau \\ s_1,s_2}} = \det[(s_1+\beta s_2)^\tau \alpha_1^{ts_1}\alpha_2^{ts_2}]_{\substack{t,\tau \\ s_1,s_2}}.$$

When we write
$$[\cdot]_{\substack{t,\tau \\ s_1,s_2}}$$
in this section, we mean an $L \times L$ matrix whose rows are indexed by pairs (t,τ) for $\tau = 0,\ldots,T_0$ and $t = -T_1,\ldots,T_1$ and whose columns are indexed by pairs (s_1,s_2) for $s_1 = -S,\ldots,S$ and $s_2 = -S,\ldots,S$. This requires an enumeration of both (t,τ) and (s_1,s_2) from $1,\ldots,L$. We fix a choice of this once and for all. Note that the sign of Δ depends on this, but this will not concern us.

Now the proof will consist of the following steps.

- Prove an analytic upper bound for Δ.
- Compute the lower bound from Liouville's inequality and comparing with the upper bound conclude $\Delta = 0$.
- Prove that $\Delta \neq 0$ and reach a contradiction.

The analytic upper bound this time will involve Schwartz's lemma from complex analysis, and we discuss it in Section 4.1. To prove $\Delta \neq 0$, we will show that certain kind of functions cannot have L zeros. This is a generalization of the fact that polynomials of degree $\leq L-1$ have this property. Such results are called zero estimates. We will discuss this in Section 4.2. We complete the proof in Section 4.3.

4.1. **Analytic upper bound.** Let $f : \mathbf{C}^n \to \mathbf{C}$ be an analytic function. By analytic, we mean that f has a power series that converges on $\mathbf{C}^n$. For $R \in \mathbf{R}_{>0}$, define
$$|f|_R = \max_{|z_1|,\ldots,|z_n| \leq R} |f(z_1,\ldots,z_n)|.$$

The purpose of this section is to prove the following upper bound on determinants.

Proposition 37. *For every $n \in \mathbf{Z}_{\geq 1}$, there is a constant $c = c(n)$ such that the following holds. Let $L \in \mathbf{Z}_{\geq 1}$ and $E \in \mathbf{R}_{>1}$. Let $f_1,\ldots,f_L : \mathbf{C}^n \to \mathbf{C}$ be analytic functions and let $\xi_1,\ldots,\xi_L \in \mathbf{C}^n$. Write*
$$r = \max_{\substack{s=1,\ldots,L \\ j=1,\ldots,n}} |\xi_{s,j}|.$$

Then
$$|\det[f_t(\xi_s)]_{\substack{t \\ s}}| \leq E^{-cL^{1+1/n}} \cdot L! \cdot \prod_{t=1}^{L} |f_t|_{Er}.$$

When we apply this result we will often take E to be a fixed constant.

Before giving the proof, we note the following corollary. Recall the determinant

$$\Delta := \det[(s_1+\beta s_2)^\tau \exp(t\lambda_1(s_1+\beta s_2))]_{\substack{t,\tau\\ s_1,s_2}}.$$

Corollary 38. *There are constants $c, C > 0$ depending only on β and λ_1 such that the following holds. Let S, T_0, T_1, L and Δ be as above. Let $E \in \mathbf{R}_{>e}$. Then*

$$|\Delta| \leq \exp(-c\log(E)L^2 + CLT_0\log(ES) + CLT_1ES).$$

Observe that the upper bound is $\leq \exp(-cL^2)$ provided $E = e$, $T_1 < c_0L^{1/2}$ and $T_0 < c_0L^{1-\varepsilon}$ with a suitably small $c_0 = c_0(\varepsilon)$. These constraints are easy to satisfy. One may even take $T_0 = c_0L$ provided E is chosen to be a small power of S.

Proof of Corollary 38. We apply the proposition with $n = 1$ for an enumeration of the functions $z^\tau \exp(t\lambda_1 z)$ in the role of f_t and for an enumeration of the points $s_1 + \beta s_2$ in the role of ξ_s. We observe that $|\xi_s| < CS$ for some constant C depending on β. In addition $|f_t(z)| < \exp(C_1T_0\log(ES) + C_1T_1ES)$ for $z \leq CES$ and some $C_1 = C_1(\beta, \lambda_1)$. We plug this into the bound in the proposition to get the result. Thanks to $E \geq e$, $\log(L!)$ will be dominated by $c\log(E)L^2$. □

We turn to the proof of the proposition. It rests on the following result from complex analysis.

Lemma 39 (Schwartz's lemma)**.** *Let $R \in \mathbf{R}_{>0}$ and let $F : B(R) \subset \mathbf{C} \to \mathbf{C}$ be a holomorphic function on the disk $B(R)$ of radius R around 0. Suppose that F vanishes to order T at 0 for some $T \in \mathbf{Z}_{\geq 0}$. Then*

$$|F(z)| \leq |F|_R \cdot \frac{|z|^T}{R^T}.$$

Proof. This follows from the maximum modulus principle applied to the function F/z^T. □

Proof of Proposition 37. We apply Schwartz's lemma for the function $F(z) = \det[f_t(z\cdot\xi_s)]_{\substack{t\\ s}}$ with $R = E$. It is immediate form the definitions that

$$|F|_E \leq L! \cdot \prod_{t=1}^{L} |f_t|_{Er}.$$

Thus the proposition follows from Schwartz's lemma applied to F at $z = 1$ if we show the following claim.
Claim. F vanishes to order at least $cL^{1+1/n}$.

We prove the claim. We first observe that it is enough to prove this in the special case, where each of the functions f_t is of the form $z_1^{a_1}\cdots z_n^{a_n}$ for some $a_1, \ldots, a_n \in \mathbf{Z}_{\geq 0}$ depending on t, of course. Indeed, we may write each f_t as a power series around $(0, \ldots, 0)$ and use the linearity of the determinant to write F as a linear combination of determinants which are in the special form described above. The order of vanishing cannot decrease by taking a linear combination, so it is indeed enough to prove the claim for determinants of the special form.

Now we assume that f_t is a monomial and write $\deg f_t$ for its total degree. We may pull out a factor of $z^{\deg f_t}$ from the t'th row for each t. This way we see that F vanishes to order at least $\sum_t \deg f_t$ at 0.

Our next observation is that we may assume that all the monomials $f_1, \ldots, f_L$ are distinct for otherwise the determinant is 0.

Now we make some crude estimates. For any $d \in \mathbf{Z}_{\geq 0}$, the number of monomials of total degree $< d$ is not more than d^n. Taking $d = \lfloor (L/2)^{1/n} \rfloor$, we see that at least $L/2$ of the monomials have total degree at least d. Therefore, the order of vanishing of F at 0 is at least $dL/2$, which is the desired bound. $\square$

4.2. **Zero estimate.**

Proposition 40. *Let $S = (T_0+1)T_1$ be positive integers. Let $\xi_1, \ldots, \xi_S \in \mathbf{R}$ and $w_1, \ldots, w_{T_1} \in \mathbf{R}$ be two sets of distinct numbers. Then*

$$\det[\xi_s^\tau e^{w_t \xi_s}]_{\substack{\tau,t\\ s}} \neq 0,$$

where the indices run through the ranges $\tau = 0, \ldots, T_0$, $t = 1, \ldots, T_1$ and $s = 1, \ldots, S$.

We deduce this from the following result.

Proposition 41. *Let $T \in \mathbf{Z}_{\geq 1}$, and let $w_1, \ldots, w_T \in \mathbf{R}$ be distinct numbers. Let $P_1, \ldots, P_T \in \mathbf{R}[x]$ be non-zero polynomials. Then the function*

$$P_1(x)e^{w_1 x} + \ldots + P_T(x)e^{w_T x}$$

has at most $\deg P_1 + \ldots + \deg P_T + T - 1$ zeros on $\mathbf{R}$ counting multiplicities.

Proof of Proposition 40 assuming Proposition 41. If the determinant is 0, we have $a_{\tau,t} \in \mathbf{R}$ for $\tau = 0, \ldots, T_0$ and $t = 1, \ldots, T_1$ such that the function

$$\sum_{\tau,t} a_{\tau,t} x^\tau e^{w_t x}$$

vanishes at $\xi_1, \ldots, \xi_S$. This function is of the form that appears in Proposition 41 with suitable choice of the polynomials $P_1, \ldots, P_T$, which are of degree at most T_0 each and there are T_1 many of them. Since $S > T_0 \cdot T_1 + T_1 - 1$, this is not possible. $\square$

The proof of Proposition 41 is based on the following calculus lemma, which is a consequence of Rolle's theorem.

Lemma 42. *Let $F : \mathbf{R} \to \mathbf{R}$ be an analytic function with n zeroes counting multiplicities. Then F' has $n-1$ zeroes.*

Proof of Proposition 41. We prove the proposition by induction on

$$N = N(T, P_1, \ldots, P_T) = \deg P_1 + \ldots + \deg P_T + T - 1.$$

If $N = 0$, then $T = 1$ and $\deg P_1 = 0$. In this case the proposition follows from the fact that $ae^{w_1 x}$ has no zeroes if $a \neq 0$.

We assume that $N > 0$ and that the claim holds for $N - 1$ in place of N. We assume, as we may that $w_1 = 0$. Indeed, if this is not the case, we may divide our function by $e^{w_1 x}$ and replace w_j by $w_j - w_1$ for all j. Now we write

$$\frac{d}{dx}(P_1(x)+P_2(x)e^{w_2 x}+\ldots+P_T(x)e^{w_T x}) = P_1'(x)+Q_2(x)e^{w_2 x}+\ldots+Q_T(x)e^{w_T x}$$

for some polynomials Q_j with $\deg Q_j = \deg P_j$ for $j = 2, \ldots, T$. In particular, $Q_j \neq 0$ for all j. We apply the induction hypothesis, which gives that

$$P_1'(x) + Q_2(x)e^{w_2 x} + \ldots + Q_T(x)e^{w_T x}$$

has at most

$$\deg P_1' + \deg Q_2 + \ldots + \deg Q_T + T - 1 = N - 1$$

zeroes. Now Lemma 42 completes the induction step. As a pedantic remark, we note that where $\deg P_1 = 0$, we apply the induction hypothesis for the function $Q_2(x)e^{w_2 x} + \ldots + Q_T(x)e^{w_T x}$ with $T - 1$ in place of T. $\square$

4.3. **Completing the proof of Theorem 13 in the real case.** Recall that $\lambda_1, \lambda_2 \neq 0$ are real numbers with $\alpha_j = e^{\lambda_j} \in \overline{\mathbf{Q}}$, and $\beta = \lambda_2/\lambda_1$ is assumed to be irrational and algebraic. Recall also the parameters L, S, T_0, T_1 satisfying $L = (T_0 + 1)(2T_1 + 1) = (2S + 1)^2$, and the determinant

$$\Delta = \det[(s_1 + \beta s_2)^\tau \exp(t\lambda_1(s_1 + \beta s_2))]_{\substack{t,\tau \\ s_1,s_2}}.$$

We apply Corollary 38, which gives

$$|\Delta| \leq \exp(-c\log(E)L^2 + CLT_0\log(ES) + CLT_1 ES).$$

Next we apply Proposition 40 for an enumeration of $(s_1 + \beta s_2)$ with $s_1, s_2 = -S, \ldots, S$ in the role of ξ_s and with $w_t = t\log\lambda_1$. Then our zero estimate gives $\Delta \neq 0$.

We also recall that

$$\Delta = \det[(s_1 + \beta s_2)^\tau \alpha_1^{ts_1}\alpha_2^{ts_2}]_{\substack{t,\tau \\ s_1,s_2}}.$$

This is an algebraic number and by Proposition 28, we have

$$H(\Delta) \leq L! \cdot 2^{LT_0} S^{LT_0} H(\beta)^{LT_0} H(\alpha_1)^{LT_1 S} H(\alpha_2)^{LT_1 S}.$$

By the Liouville bound, we get

$$|\Delta| \geq \exp(-C(L\log L + LT_0\log S + LT_1 S))$$

for some constant C depending on $\alpha_1, \alpha_2, \beta$. This contradicts our previous upper bound, provided we choose the parameters in such a way that $E \geq e$, $L = (T_0 + 1)(2T_1 + 1) = (2S + 1)^2$ and

$$\log(E)L > C(T_0 \log(ES) + T_1 ES)$$

for some C. As we already indicated after Corollary 38, there are plenty of ways to do this.

4.4. **Zero estimate – complex case.** The general case of the Gelfond–Schneider theorem requires a more refined zero estimate. This is also needed if we want to show lower bounds on the linear form in logarithms $\beta\lambda_1 - \lambda_2$ as opposed to just non-vanishing.

Theorem 43 (Zero estimate, Nesterenko)**.** *Let T_0, T_1, N and M be positive integers, and let $\Sigma_1, \Sigma_2 \subset \mathbf{C}^2$ with $|\Sigma_1| = N$ and $|\Sigma_2| = M$ and such that the exponentials of the second coordinates of Σ_1 and the first coordinates of Σ_2 are distinct. Let $P \in \mathbf{C}[X, Y]$ be a non-zero polynomial of degree at most T_0 in X and at most T_1 in Y such that the function $P(X, \exp(y))$ vanishes at all points of $\Sigma_1 + \Sigma_2$. Then either*

$$N \leq T_1 \quad \text{or} \quad M \leq T_0(T_1 + 1).$$

The strategy of the proof is the following. We assume, as we may that P is not divisible by Y. If it were, then writing $P(X, Y) = \widetilde{P}(X, Y)Y$, we observe that $P(X, \exp(y))$ and $\widetilde{P}(X, \exp(y))$ vanish on precisely the same points, since $\exp(y)$ is never 0.

We suppose that $N > T_1$, and write

$$\Sigma_1 = \{(\xi_1, \eta_1), \ldots, (\xi_N, \eta_N)\}.$$

We observe that the functions $P(\xi_j + X, \exp(\eta_j + y))$ for $j = 1, \ldots, N$ all vanish on Σ_2. We aim to eliminate the variable y and construct a polynomial in X alone that vanishes on Σ_2. More concretely, we look for polynomials $A_1(X), \ldots, A_N(X)$ such that

$$\begin{aligned} A_1(X)P(\xi_1 + X, \exp(\eta_1 + y)) + \ldots \\ + A_N(X)P(\xi_N + X, \exp(\eta_N + y)) = B(X) \end{aligned}$$

for some non-zero polynomial B. We will aim to do this in such a way that $\deg B \leq T_0(T_1 + 1)$. Then B may have at most $T_0(T_1 + 1)$ distinct roots, hence $M \leq T_0(T_1 + 1)$ will follow.

Our first lemma allows us to eliminate the Y variable from a collection of polynomials of the form $Q_{i,1}(X) + Q_{i,2}Y^{n_2} + \ldots + Q_{i,K}(X)Y^{n_K}$ for $i = 1, \ldots, K$.

Lemma 44. *Let $Q_{i,j} \in \mathbf{C}[X]$ be polynomials for $i, j = 1, \ldots, K$ and let $B = \det[Q_{i,j}]_{i,j}$. Then there are polynomials $A_1, \ldots, A_K \in \mathbf{C}[X]$ such that*

$$(A_1, \ldots, A_K)[Q_{i,j}]_{i,j} = (B, 0, \ldots, 0).$$

Proof. Write $[Q^{\mathrm{adj}}_{i,j}]_{i,j}$ for the adjugate matrix of $[Q_{i,j}]_{i,j}$. Then

$$[Q^{\mathrm{adj}}_{i,j}]_{i,j}[Q_{i,j}]_{i,j} = \det[Q_{i,j}]_{i,j} \cdot \mathrm{Id}\,.$$

So we may simply take $A_j = Q^{\mathrm{adj}}_{1,j}$. ☐

Let

$$P(X,Y) = R_1(X)Y^{n_1} + \ldots + R_K(X)Y^{n_K}$$

such that $R_1, \ldots, R_K \in \mathbf{C}(X)$ are non-zero polynomials and $0 = n_1 < n_2 < \ldots < n_K \le T_1$. (Recall that Y does not divide P.) Note that $K \le T_1 + 1$. Observe that

$$P(\xi+X, \exp(\eta+y)) = R_1(\xi+X)\exp(n_1\eta)Y^{n_1}+\ldots+R_K(\xi+X)\exp(n_K\eta)Y^{n_K},$$

where $Y = \exp(y)$.

We will use Lemma 44 with

$$Q_{i,j}(X) = R_j(\xi_i + X)\exp(n_j\eta_i),$$

where (ξ_i, η_i) are appropriately chosen elements of Σ_1 for $i = 1, \ldots, K$. We need to show that for an appropriate choice, $\det[Q_{i,j}]_{i,j}$ is non-zero, otherwise the lemma is not very useful. When one needs to show that a polynomial is non-zero, it is often a good strategy to show that its leading coefficient is not 0. Observe that the leading coefficient of $Q_{i,j}$ is $a_j \exp(n_j\eta_i)$, where $a_j \ne 0$ is the leading coefficient of R_j. Now the leading coefficient of a determinant with polynomial entries is the determinant of the leading coefficients of the entries provided this is not zero. Therefore, the leading coefficient of $\det[Q_{i,j}]_{i,j}$ is

$$a_1 \cdots a_K \det[\exp(n_j\eta_i)],$$

provided this is not 0.

Lemma 45. *Let $K \in \mathbf{Z}_{\ge 1}$ and let $0 \le n_1 < \ldots < n_K$. Let $A \subset \mathbf{C}$ such that $|\{\exp(\eta) : \eta \in A\}| > n_K$. Then, there is a choice of $\eta_1, \ldots, \eta_K \in A$ such that*

$$\det[\exp(n_j\eta_i)] \ne 0.$$

Proof. The proof is by induction on K. If $K = 1$, the lemma is trivial. We assume $K > 1$ and that the claim holds for $K - 1$. Let $\eta_1, \ldots, \eta_{K-1} \in A$ be such that

$$\det[\exp(n_j\eta_i)]_{\substack{i=1,\ldots,K-1\\ j=1,\ldots,K-1}} \ne 0. \tag{26}$$

Now consider the polynomial

$$F(Z) = \begin{vmatrix} \exp(n_1\eta_1) & \ldots & \exp(n_K\eta_1) \\ \vdots & \ddots & \vdots \\ \exp(n_1\eta_{K-1}) & \ldots & \exp(n_K\eta_{K-1}) \\ Z^{n_1} & \ldots & Z^{n_K} \end{vmatrix}.$$

Observe that the coefficient of Z^{n_K} in F is (26), which is not zero. For this reason, F is a non-zero polynomial of degree n_K, hence it has at most n_K roots. We can choose η_K in such a way that $\exp(\eta_K)$ is not among the roots of F, and this completes the proof. $\square$

Now we complete the proof of Theorem 43 in a few strokes. By Lemma 45, there is a choice of $(\xi_1, \eta_1), \ldots, (\xi_K, \eta_K) \in \Sigma_1$ such that for

$$Q_{i,j}(X) = R_j(\xi_i + X)\exp(n_j\eta_i),$$

we have $B = \det[Q_{i,j}]_{i,j} \neq 0$, where R_j are as above. Observe that $\deg B \leq T_0 K \leq T_0(T_1+1)$. Note that

$$\begin{pmatrix} Q_{1,1} & \ldots & Q_{1,K} \\ \vdots & \ddots & \vdots \\ Q_{K,1} & \ldots & Q_{K,K} \end{pmatrix} \begin{pmatrix} Y^{n_1} \\ \vdots \\ Y^{n_K} \end{pmatrix} = \begin{pmatrix} P(\xi_1 + X, \exp(\eta_1 + y)) \\ \vdots \\ P(\xi_K + X, \exp(\eta_K + y)) \end{pmatrix}.$$

Recall that $n_1 = 0$. By Lemma 44, there are $A_1, \ldots, A_K$ such that

$$(A_1(X), \ldots, A_K(X)) \begin{pmatrix} P(\xi_1 + X, \exp(\eta_1 + y)) \\ \vdots \\ P(\xi_K + X, \exp(\eta_K + y)) \end{pmatrix} = B(X).$$

Now B is a non-zero polynomial of degree at most $T_0(T_1+1)$ that vanishes at the first coordinates of the points in Σ_2. It follows that $M = |\Sigma_2| \leq T_0(T_1+1)$, as required.

4.5. **Gelfond–Schneider - general case (Non-examinable).** In this section we briefly discuss how to modify the proof of the Gelfond Schneider theorem we gave above to work in the general complex case.

Using Nesterenko's zero estimate, Theorem 43, in place of Proposition 41, we may prove the following variant of Proposition 40.

Proposition 46. *Let T_0, T_1, and S be positive integers such that*

$$S > T_1, \qquad S^2 > T_0(T_1+1).$$

Let $\beta, \lambda_1, \lambda_2 \in \mathbf{C}$ such that $\beta \notin \mathbf{Q}$ and at least one of λ_1 or λ_2 is not in $i\pi\mathbf{Q}$. Consider the $(T_0+1)(T_1+1) \times (2S+1)^2$ matrix

$$M = [(s_1 + \beta s_2)^\tau \exp(t(\lambda_1 s_1 + \lambda_2 s_2))]_{\substack{\tau,t \\ s_1,s_2}}$$

with the indices running through the ranges $\tau = 0, \ldots, T_0$, $t = 0, \ldots, T_1$ and $s_1, s_2 = -S, \ldots, S$. Then

$$\operatorname{rank} M = (T_0+1)(T_1+1).$$

We sketch the proof. Multiplying M by a row vector from the left, we get a row vector, which gives the values of a function of the form $P(X, \exp(y))$ evaluated at the points $(s_1 + \beta s_2, \lambda_1 s_1 + \lambda_2 s_2)$, where P is a polynomial of degree at most T_0 in X and at most T_1 in $\exp(y)$. To prove the proposition, we need to show that this row cannot be the 0 vector. To this end, we apply Theorem 43 with

$$\Sigma_1 = \{(s, \lambda_1 s) : |s| \le S/2\},$$
$$\Sigma_2 = \{(s_1 + \beta s_2, \lambda_1 s_1 + \lambda_2 s_2) : |s_1|, |s_2| \le S/2\}$$

provided $\lambda_1 \notin i\pi\mathbf{Q}$. In the other case, we modify Σ_1 accordingly. Now observe that $|\Sigma_1| \ge S > T_1$ and $|\Sigma_2| \ge S^2 > T_0(T_1)$, and in the respective sets the exponentials of the second coordinates and the first coordinates are distinct. In addition, the elements of $\Sigma_1 + \Sigma_2$ are of the form $(s_1 + \beta s_2, \lambda_1 s_1 + \lambda_2 s_2)$ with $|s_1|, |s_2| \le S$, so Theorem 43 applies and implies the claim.

Now we sketch how Proposition 46 may be used in place of Proposition 40 to prove the general case of the Gelfond–Schneider theorem. Let $\lambda_1 = \log(\alpha_1)$ and $\lambda_2 = \log(\alpha_2)$ be logarithms of algebraic numbers, and suppose to the contrary that $\beta = \lambda_2/\lambda_1$ is an irrational algebraic number. We begin by extracting a $(T_0 + 1)(T_1 + 1) \times (T_0 + 1)(T_1 + 1)$ determinant Δ from the matrix M, which does not vanish. Observe that Δ is an algebraic number and we may estimate its height in the same way as we did for the corresponding determinant in Section 4.3.

Now we exploit that $\beta = \lambda_2/\lambda_1$. This allows us to write Δ as an alternant of a function of a single variable. As we have seen in Section 4.1, this is crucial to get a good enough upper bound. We may write

$$(s_1 + \beta s_2)^\tau \exp(t(\lambda_1 s_1 + \lambda_2 s_2)) = \varphi_{\tau,t}(s_1 + \beta s_2)$$

with $\varphi_{\tau,t}(x) = x^\tau \exp(t\lambda_1 x)$. We can then use Proposition 37 to prove an upper bound on $|\Delta|$ similar to Corollary 38, which contradicts the Liouville bound, and this completes the proof.

4.6. **Lower bounds for linear forms in logarithms (Non-examinable).** In this section, we sketch out how the ideas in the previous sections can be used to give a lower bound for the linear form in logarithms

$$\Lambda = \lambda_2 - \beta\lambda_1,$$

where $\lambda_j = \log(\alpha_j)$.

We consider two determinants. The first one is the "arithmetic" one

$$\Delta_{\mathrm{ar}} = \det[(s_1^{(j)} + \beta s_2^{(j)})^\tau \alpha_1^{ts_1^{(j)}} \alpha_2^{ts_2^{(j)}}]_{\tau,t,j},$$

where $\tau = 0, \ldots, T_0$, $t = 0, \ldots, T_1$ and $(s_1^{(j)}, s_2^{(j)})$ are appropriate elements of $\{-S, \ldots, S\}^2$ for $j = 1, \ldots, (T_0 + 1)(T_1 + 1)$ chosen using Proposition 46 so that $\Delta_{\mathrm{ar}} \ne 0$.

The second determinant is the "analytic" one

$$\Delta_{\mathrm{an}} = \det[(s_1^{(j)} + \beta s_2^{(j)})^\tau \exp(t\lambda_1(s_1^{(j)} + \beta s_2^{(j)}))]_{\tau,t,j}.$$

Just as before, we can get a lower bound on $\Delta_{\rm ar}$ by Liouville's inequality and an upper bound on $\Delta_{\rm an}$ using Schwartz's lemma. What is different now is that the two determinant may not equal. The entries of the two determinants are related to each other by the formula

$$(s_1^{(j)}+\beta s_2^{(j)})^\tau \alpha_1^{ts_1^{(j)}} \alpha_2^{ts_2^{(j)}} = (s_1^{(j)}+\beta s_2^{(j)})^\tau \exp(t\lambda_1(s_1^{(j)}+\beta s_2^{(j)})) \exp(t\Lambda s_2^{(j)}).$$

Provided Λ is sufficiently small, it is possible to derive a contradiction between the upper and lower bounds. This can be turned into an indirect argument to show a lower bound on Λ.

5. Integral point on curves

The purpose of this section is to prove the following result using the subspace theorem.

Theorem 47. *Let $d \in \mathbf{Z}_{\geq 3}$. Let $F \in \mathbf{Z}[X,Y]$ be a homogeneous polynomial of degree d without repeated factors. Let $G \in \mathbf{Z}[X,Y]$ be a polynomial of degree at most $d-1$. Suppose that $F-G$ is irreducible in $\mathbf{Z}[X,Y]$. Then the equation*

$$F(X,Y) = G(X,Y), \qquad X,Y \in \mathbf{Z}$$

has at most finitely many solutions.

Where $F-G$ is not irreducible, we can factorize it, and try to apply the theorem to the irreducible factors.

The above theorem is a weaker form of a result of Schnizel, who proved the same result only assuming that F is not of the form aQ^n for some $a \in \mathbf{Z}$ and an irreducible polynomial Q of degree at most 2 instead of that it has no repeated factors. Schinzel proved his theorem using a famous result of Siegel, which states that if an affine algebraic curve has infinitely many integral points, then it must have genus 0 and at most 2 points at infinity. Here we will prove Theorem 47 using the subspace theorem. The proof is based on an approach of Corvaja and Zannier for proving Siegel's theorem. Finally, we mention a deep result of Faltings which states that an algebraic curve with infinitely many rational points must have genus 0 or 1.

We prove the above theorem using the subspace theorem, which we restate now in an equivalent form with different notation.

Theorem 48 (Subspace theorem)**.** *Let V be an $n \in \mathbf{Z}_{\geq 2}$ dimensional vector space over $\overline{\mathbf{Q}}$. Let $\ell_1,\ldots,\ell_n \in V$ and $\ell_1^{(0)},\ldots,\ell_n^{(0)} \in V$ be two bases. Then for all $\varepsilon > 0$, there is a finite collection of elements $f_1,\ldots,f_m \in V_{\neq 0}$ such that all elements $\varphi \in V^*$ of the dual space solving* (27)

$$\prod_{j=1}^n |\varphi(\ell_j)| \leq H(\varphi(\ell_1^{(0)}),\ldots,\varphi(\ell_n^{(0)}))^{-\varepsilon}, \qquad \varphi(\ell_1^{(0)}),\ldots,\varphi(\ell_n^{(0)}) \in \mathbf{Z}$$

satisfy $\varphi(f_j) = 0$ *for some* $j = 1, \dots, m$.

We begin by explaining the relationship between this formulation and Theorem 6. We have some $\alpha_{i,j} \in \overline{\mathbf{Q}}$ such that

$$\ell_j = \alpha_{1,j}\ell_1^{(0)} + \ldots + \alpha_{n,j}\ell_n^{(0)}$$

and we use these to construct linear forms

$$L_j = \alpha_{1,j}X_1 + \ldots + \alpha_{n,j}X_n.$$

Since (ℓ_j) is a basis, these are linearly independent. Now the map $\varphi \mapsto (\varphi(\ell_1^{(0)}), \dots, \varphi(\ell_n^{(0)}))$ is an isomorphism between V^* and $\overline{\mathbf{Q}}^n$ and we have

$$L_j(\varphi(\ell_1^{(0)}), \dots, \varphi(\ell_n^{(0)})) = \varphi(\ell_j).$$

Therefore, $\varphi \in V^*$ satisfies (27) if and only if $(x_1, \dots, x_n) = \varphi(\ell_1^{(0)}), \dots, \varphi(\ell_n^{(0)})$ satisfies

$$\prod_{j=1}^{n} |L_j(x_1, \dots, x_n)| \le H(x_1, \dots, x_n)^{-\varepsilon}, \qquad (x_1, \dots, x_n) \in \mathbf{Z}.$$

This shows that Theorem 48 is completely equivalent to 6, and only the notation is different. In many applications of the subspace theorem, the choice of the "reference basis" $\ell_1^{(0)}, \dots, \ell_n^{(0)}$ is completely irrelevant so long as we do not change the $\mathbf{Q}$-vector space it generates. It is often more convenient not to introduce coordinates and use the theorem in the above form. This is especially the case when we apply the subspace theorem repeatedly in different vector spaces, which we will do in later sections.

Now we turn to the proof of Theorem 47. We let F and G be as in the theorem and write $P = F - G$. For simplicity, we assume that $Y \nmid F$. (This can be achieved by a suitable linear change of variables if necessary.) Then there are some distinct $\alpha_1, \dots, \alpha_d \in \overline{\mathbf{Q}}$ and $a \in \mathbf{Z}$ such that

$$F(X, Y) = a(X - \alpha_1 Y) \cdots (X - \alpha_d Y).$$

Let $x, y \in \mathbf{C}$ with $P(x, y) = 0$. Then

$$F(x, y) \le C(|x| + |y|)^{d-1}$$

for some constant $C = C(P)$. By an argument similar to Lemma 5, for every $\varepsilon > 0$, there is a constant $R = R(P, \varepsilon) \in \mathbf{R}_{>0}$ such that $|x| + |y| > R$, we have $|x/y - \alpha_j| < \varepsilon$ for some $j = 1, \dots, d$.

We fix a suitably small ε assuring, in particular, that $|\alpha_i - \alpha_j| > 2\varepsilon$ for any pairs of indices $i \neq j$. We write

$$\Gamma_0 = \{(x, y) \in \mathbf{C}^2 : P(x, y) = 0, |x| + |y| \le R\}$$
$$\Gamma_j = \{(x, y) \in \mathbf{C}^2 : P(x, y) = 0, |x| + |y| > R, |x/y - \alpha_j| < \varepsilon\}$$

for $j = 1, \dots, d$. It follows that $\Gamma_0, \Gamma_1, \dots, \Gamma_d$ is a partition of the solution set Γ of the equation $P(x, y) = 0$. It is easy to show that for all $y \in \mathbf{C}$ sufficiently large in modulus, there is a unique $x \in \mathbf{C}$ such that $(x, y \in \Gamma_j)$, but we do not need to know this.

Since Γ_0 is bounded, it may contain at most finitely many integer points. We are going to prove the same for Γ_j for $j > 0$. Let $I = P\overline{\mathbf{Q}}[X, Y]$ be the ideal generated by P. We fix a suitably large number D and consider the vector space $\overline{\mathbf{Q}}[X, Y]^{(D)}$ of polynomials of total degree at most D. We will apply the subspace theorem in the vector space

$$V = \overline{\mathbf{Q}}[X, Y]^{(D)}/(\overline{\mathbf{Q}}[X, Y]^{(D)} \cap I).$$

Note that elements of V may be evaluated on points in Γ. Indeed, elements of V are polynomials in $\overline{\mathbf{Q}}[X, Y]$ that are determined up to adding an element of I. However, the elements of I vanish identically on Γ, so each representative in a coset of I gives the same value to points in Γ. Moreover, for any fixed $(x, y) \in \Gamma \cap \overline{\mathbf{Q}}^2$, the map

$$V \to \overline{\mathbf{Q}} : \ f \mapsto f(x, y)$$

is an element of the dual V^*.

The images of the monomials X^iY^j for $i, j \in \mathbf{Z}_{\geq 0}$ with $i + j \leq D$ span V. We select the reference basis $\ell_1^{(0)}, \ldots, \ell_n^{(0)} \in V$ with $n = \dim V$ from them. We fix some choice of this, it is not important which one. If $(x, y) \in \Gamma \cap \mathbf{Z}^2$ is an integer point, then $\ell_j^{(0)}(x, y) \in \mathbf{Z}$ for $j = 1, \ldots, n$ and

$$H(\ell_1^{(0)}(x, y), \ldots, \ell_n^{(0)}(x, y)) \leq C|y|^D$$

for some constant C. Here we used that x/y is close to α_j for some j, so x can be bounded in terms of y. We are going to use this repeatedly without mentioning it again.

The proof hinges on our ability to construct elements of V that take small values on Γ_j for some fixed $j > 0$. We are now going to discuss this. Before we begin, we warn that readers familiar with basic algebraic geometry may find what follows a bit silly. This is because algebraic geometry is not a prerequisite for this course. However, we are going to make some non-examinable remarks to explain the link with more conventional terminology. These will also be suggestive towards extending the argument for proving Siegel's theorem on integral points on curves.

For each $j = 1, \ldots, d$, we introduce a symbol p_j, and call them the points of Γ at infinity. We define the order of $f \in V$ at p_j as

$$\mathrm{ord}_{p_j}(f) = \sup(t \in \mathbf{Z} : f(X, Y) \cdot Y^t \text{ is bounded on } \Gamma_j).$$

Note that always $|f(X, Y)| \leq C|Y|^D$ on Γ_j for all $f \in V$ and some constant $C = C(f)$. Hence $\mathrm{ord}_{p_j}(f) \geq -D$ for all j and $f \in V$.

Lemma 49. *Let $f \in V$ and $j \in \{1, \ldots, d\}$. If $\mathrm{ord}_{p_j}(f) < \infty$, then the limit*

$$\lim_{(X,Y)\in\Gamma_j:|Y|\to\infty} f(X, Y) \cdot Y^{\mathrm{ord}_{p_j}(f)}$$

exists and is non-zero.

In addition, we have

$$\lim_{(X,Y)\in\Gamma_j:|Y|\to\infty}(X-\alpha Y)\cdot Y^{-1}=\alpha_j-\alpha$$

for all $\alpha\in\overline{\mathbf{Q}}$.

It can be proved that $\mathrm{ord}_{p_j}(f)=\infty$ if and only if $f=0$, but we do not need to know this.

Remark 50 (Non-examinable)**.** Multiplying each monomial of P by a suitable power of Z to make it homogeneous, we may consider Γ as a projective curve embedded in $\mathbf{P}^2$. Then Γ has indeed d points on the line $Z=0$ given by $p_j=(\alpha_j:1:0)$ for $j=1,\dots,d$.

Given $f\in V$, we may divide each monomial of f by a power of Z to make it a rational function on the projective curve Γ. Observe that this does not alter f on the affine chart $\{(X:Y:1)\}$. Note that the rational function Y/Z is equal to Y on the chart $\{(X:Y:1)\}$. The lemma should be interpreted as saying that Z/Y is a local uniformizer at each p_j.

Proof. To simplify notation, we take $j=1$. Using the substitution $(X-\alpha_1Y)\mapsto X$, we may reduce to the case $\alpha_1=0$. We first show that X is bounded on Γ_1. To this end, observe that $F(X,Y)=G(X,Y)$ implies

$$X=\frac{G(X,Y)}{(X-\alpha_2Y)\cdots(X-\alpha_dY)}$$

on Γ. On Γ_1, the denominator is bounded below by $c|Y|^{d-1}$ for some constant $c>0$. The numerator has degree $d-1$, so the claim follows.

This immediately implies

$$\lim_{(X,Y)\in\Gamma_1:|Y|\to\infty}(X-\alpha Y)\cdot Y^{-1}=-\alpha,$$

and the last claim of the lemma follows.

Next we observe that $P(X,Y)=aXY^{d-1}+bY^{d-1}+\widetilde{P}(X,Y)$, where $a,b\in\mathbf{Z}$ and the degree of $\widetilde{P}(X,Y)$ in Y is at most $d-2$. Moreover, $a\neq 0$ because $X^2\nmid F$. Then

$$X=-\frac{b}{a}+Y^{-1}R(X,Y^{-1})$$

on Γ for some polynomial R. In particular, $\lim_{|Y|\to\infty}X=-\frac{b}{a}$ on Γ_1.

Suppose $f\in V$ can be written in the form

$$f(X,Y)=Q(X)Y^k+S(X,Y^{-1})Y^{k-1} \tag{28}$$

for some $k\in\mathbf{Z}$ and polynomials Q and S. This can always be achieved for $k=D$. Now we see that

$$\lim_{|Y|\to\infty}f(X,Y)\cdot Y^{-k}=Q(-b/a).$$

If $Q(-b/a) \neq 0$, the claim is proved. If $Q(-b/a) = 0$, then

$$Q(X) = Q(-\frac{b}{a} + Y^{-1}R(X, Y^{-1})) = Y^{-1}T(X, Y^{-1})$$

for some polynomial T, and we can write

$$f(X, Y) = T(X, Y^{-1})Y^{k-1} + S(X, Y^{-1})Y^{k-1}$$

in the form (28) with $k - 1$ in place of k.

We iterate this procedure. Either it terminates, and the claim is proved, or we get that $f(X, Y)Y^t$ is bounded for all $t \in \mathbf{Z}$ on Γ_1 and $\mathrm{ord}_{p_1}(f) = \infty$.

□

The next lemma constructs the basis of V that will be used in the subspace theorem.

Lemma 51. *For each $j \in \{0, \ldots, d\}$, there is a basis $\ell_1, \ldots, \ell_n \in V$ such that*

$$\mathrm{ord}_{p_j}(\ell_i) \geq -D + i - 1$$

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

Proof. We prove by induction the following statement. For all $i = 0, \ldots, n$, there are $\ell_1, \ldots, \ell_i$ and a subspace $V_i \subset V$ of dimension $n - i$ such that

$$\begin{aligned} V &= \overline{\mathbf{Q}}\ell_1 \oplus \ldots \oplus \overline{\mathbf{Q}}\ell_i \oplus V_i, \\ \mathrm{ord}_{p_j}(\ell_k) &\geq -D + k - 1 \qquad \text{for } k = 1, \ldots, i, \\ \mathrm{ord}_{p_j}(f) &\geq -D + i \qquad \text{for all } f \in V_i. \end{aligned}$$

The lemma is the $i = n$ case of this.

The claim is trivial for $i = 0$ with $V_0 = V$. Suppose $i > 0$ and that the claim holds for $i - 1$. We take any element of V_{i-1} with minimal order as ℓ_i and let

$$V_i = \{f \in V_{i-1} : \mathrm{ord}_{p_j}(f) > \mathrm{ord}_{p_j}(\ell_i)\}.$$

We show

$$V_{i-1} = \overline{\mathbf{Q}}\ell_i \oplus V_i.$$

To this end, let $m = \mathrm{ord}_{p_j}(\ell_i)$ and

$$\alpha = \lim_{(X,Y) \in \Gamma_j, |Y| \to \infty} \ell_i(X, Y)Y^m.$$

By Lemma 49, $\alpha \neq 0$. Now take any $f \in V_{i-1}$ and let

$$\beta = \lim_{(X,Y) \in \Gamma_j, |Y| \to \infty} f(X, Y)Y^m.$$

Taking $g = f - (\beta/\alpha)\ell_i$, we get

$$\lim_{(X,Y) \in \Gamma_j, |Y| \to \infty} g(X, Y)Y^m.$$

By Lemma 49, this implies $\mathrm{ord}_{p_j}(g) > m$, so $g \in V_i$, which proves the claim.

This completes the induction and the proof of the lemma. $\square$

The first few elements of the basis constructed above grow rapidly with $|Y|$ on Γ_j. However, if the dimension of V is sufficiently large, then the later elements in the basis will decay. We will see that for the application to work, we need $\dim V \geq 2D + 2$. This is achieved by the next lemma.

Lemma 52. *We have* $\dim V \geq dD - d(d-1)$.

Remark 53 (Non-examinable)**.** Note that V is the space of rational functions on Γ, which have a pole of order at most D at each of the points at infinity and no other poles. Using this as the definition of V in the case of a general curve, one may prove Siegel's theorem using the subspace theorem by the argument we present for plane curves. We add that by the Riemann-Roch theorem, we have $\dim V = dD + 1 - g$ at least when $dD \geq 2g - 1$, where g is the genus of the curve. The above weaker bound is sufficient for our purposes.

Proof. Let

$$R(X,Y) = \prod_{j=1}^{d}(X - \alpha_j Y)$$

We will show that the polynomials

$$Q_{j,l}(X,Y) = R(X,Y)(X - \alpha_j Y)^{-1} Y^l \in V$$

for $j = 1, \ldots, d$ and $l = 1, \ldots D - d + 1$ are linearly independent. This proves $\dim V \geq d(D - d + 1)$, as needed.

Let

$$Q = \sum_{j,l} \beta_{j,l} Q_{j,l}$$

such that not all $\beta_{j,l}$ are 0. We show that $Q \neq 0$ in V. Suppose that $\beta_{j',l'} \neq 0$ and that l' is maximal with respect to this property. We show that

$$\lim_{(X,Y)\in\Gamma_{j'}:|Y|\to\infty} Q(X,Y) \cdot Y^{-l'-d+1} = \beta_{j',l'} \prod_{i\in\{1,\ldots,d\}\setminus\{j'\}} (\alpha_{j'} - \alpha_i) \neq 0,$$

which proves $Q \neq 0$.

By Lemma 49, we have

$$\lim_{(X,Y)\in\Gamma_{j'}:|Y|\to\infty} Q_{j',l'}(X,Y) \cdot Y^{-l'-d+1} = \prod_{i\in\{1,\ldots,d\}\setminus\{j'\}} (\alpha_{j'} - \alpha_i),$$

and

$$\lim_{(X,Y)\in\Gamma_{j'}:|Y|\to\infty} Q_{j,l}(X,Y) \cdot Y^{-l'-d+1} = 0$$

for all j, l with $l \leq l'$ unless $l = l'$ and $j = j'$. This proves the claim. $\square$

We will prove Theorem 47 applying the subspace theorem for the basis of V constructed in Lemma 51. From this we will learn that there are finitely many $f_1, \ldots, f_m \in V$ such that one of them must vanish on each integer point of Γ_j. Then we will need to following lemma to finish the proof.

Lemma 54. *Let $P, f \in \overline{\mathbf{Q}}[X,Y]$ be non-zero polynomials without common factors in $\overline{\mathbf{Q}}[X,Y]$. Then the system of equations $P[X,Y] = f[X,Y] = 0$ have finitely many solutions in $\overline{\mathbf{Q}}^2$.*

Proof. We may consider P and f to be polynomials in Y with coefficients in $\overline{\mathbf{Q}}[X]$, and we write $\overline{\mathbf{Q}}[X][Y]$ for this ring. This is canonically isomorphic to $\overline{\mathbf{Q}}[X,Y]$, so P and f have no common factors in $\overline{\mathbf{Q}}[X][Y]$. By a version of Gauss's lemma, this implies that P and f have no common factors in the ring $\overline{\mathbf{Q}}(X)[Y]$, either. This is because $\overline{\mathbf{Q}}[X]$ is a UFD and $\overline{\mathbf{Q}}(X)$ is its quotient field.

Now $\overline{\mathbf{Q}}(X)[Y]$ is a Euclidean domain, so we may find $F, G \in \overline{\mathbf{Q}}(X)[Y]$ such that

$$FP + Gf = 1.$$

Multiplying by the denominators of the $\overline{\mathbf{Q}}(X)$ coefficients of F and G, we find $\widetilde{F}, \widetilde{G} \in \overline{\mathbf{Q}}[X][Y]$ and $D \in \overline{\mathbf{Q}}[X]$ such that

$$\widetilde{F}(X,Y)P(X,Y) + \widetilde{G}(X,Y)f(X,Y) = D(X).$$

Therefore, the X coordinate of any common zero of P and f must be a zero of D. However, there are only finitely many such zeros, so we see that the X coordinates of the common zeros lie in a finite set.

Exchanging X and Y, we can see the same about the Y coordinates. This shows that the set of common zeros must be finite. $\square$

Now we can conclude the proof of the theorem.

Proof of Theorem 47. We fix some $j = 1, \ldots, d$ and show that there are only finitely many integer points in Γ_j.

We will apply the subspace theorem in the form of Theorem 48. We take $V = \overline{\mathbf{Q}}[X,Y]^{(D)}/(I \cap \overline{\mathbf{Q}}[X,Y]^{(D)})$, write $n = \dim V$ and we let $\ell_1^{(0)}, \ldots, \ell_n^{(0)} \in V$ be an appropriate collection of monomials as explained above. We use Lemma 51 to find a basis $\ell_1, \ldots, \ell_n$ such that $\mathrm{ord}_{p_j}(\ell_i) \geq -D + i - 1$ for all $i = 1, \ldots, n$. Then we have

$$\prod_{i=1}^{n} |\ell_i(X,Y)| \leq C \prod_{i=1}^{n} |Y|^{D-i+1}$$

on Γ_j with some constant C that depends on $\ell_1, \ldots, \ell_n$ so ultimately only on P, D.

By Lemma 52, if D is sufficiently large, $n \geq 2D+2$. Then

$$\prod_{i=1}^{n} |\ell_i(X,Y)| \leq C|Y|^{-D-1},$$

where C is a constant depending only on P and D, potentially different from the previous one.

As we have already noted,

$$H(\ell_1^{(0)}(x,y),\ldots,\ell_n^{(0)}(x,y)) \leq C|y|^D$$

for $(x,y) \in \Gamma_j \cap \mathbf{Z}^2$ for yet another constant C, and we get

$$\prod_{i=1}^{n} |\ell_i(x,y)| \leq H(\ell_1^{(0)}(x,y),\ldots,\ell_n^{(0)}(x,y))^{-1}$$

provided $|y|$ is sufficiently large.

Now the subspace theorem implies that there are finitely many non-zero polynomials $f_1,\ldots,f_m \in V$ such that the points $(x,y) \in \Gamma_j \cap \mathbf{Z}^2$ with sufficiently large $|y|$ must satisfy $f_i(x,y)$ for some i.

If P has a no common factor with any of the f_i, then the proof is complete by Lemma 54.

In the other case, P must be reducible over $\overline{\mathbf{Q}}$, because the $f_i \neq 0$ in V, so $P \nmid f_i$ in $\overline{\mathbf{Q}}[X,Y]$. Let $P = aQ_1\cdots Q_k$ be the factorization of P to irreducible factors in $\overline{\mathbf{Q}}[X,Y]$ such that each Q_i is monic. Suppose that $Q_1,\ldots,Q_l$ are all the distinct Galois conjugates of Q_1 (with respect to the Galois group of a Galois extension of $\mathbf{Q}$ containing all the coefficients of all the Q_i). Then $Q_1\cdots Q_l \in \mathbf{Q}[X,Y]$, being invariant under the Galois group. Since P is irreducible in $\mathbf{Q}[X,Y]$, we must have $l=k$, and all the Q_i are Galois conjugates of each other and distinct.

Now suppose $x,y \in \mathbf{Z}$ and $P(x,y)=0$. Then $Q_i(x,y)=0$ for some i. However, (x,y) is fixed by the Galois group, and all the Q_i are Galois conjugates, so it follows that $Q_i(x,y)=0$ for all i. We apply Lemma 54 again, which implies that the set of common zeros of all the Q_i is finite, so the proof is complete. □

6. The order $\times 2, \times 3$ in $\mathbf{Z}/q\mathbf{Z}$

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

Theorem 55. *For an integer $q \in \mathbf{Z}_{\geq 1}$ with $\gcd(q,6)=1$, write $\operatorname{ord}(q)$ for the order of the multiplicative subgroup of $\mathbf{Z}/q\mathbf{Z}$ generated by 2 and 3. Then*

$$\lim_{q\to\infty} \frac{\operatorname{ord}(q)}{(\log q)^2} = \infty.$$

This follows very easily from the following result, which was proved independently by Corvaja and Zannier [6], and Hernández and Luca [7] after initial work by Bugeaud, Corvaja and Zannier [4].

Theorem 56 (Corvaja, Zannier; Hernández, Luca)**.** *Write $\mathcal{S}$ for the set of numbers of the form 2^n3^m for $n, m \in \mathbf{Z}_{\geq 0}$. Then for all $\varepsilon > 0$, there are only finitely many pairs of multiplicatively independent $a, b \in \mathcal{S}$ such that*

$$\gcd(a-1, b-1) \geq \max(a, b)^{\varepsilon}.$$

Two integers a, b are multiplicatively independent if there are no non-zero $n, m \in \mathbf{Z}$ with $a^n = b^m$.

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

$$\gcd(2^n - 1, 3^n - 1) \geq 3^{n^{c/\log\log n}}.$$

Proof of Theorem 55. Let

$$\Lambda = \{(n, m) \in \mathbf{Z}^2 : 2^n3^m \equiv 1 \mod q\}.$$

Then Λ is a lattice inside $\mathbf{Z}^2$ and $[\mathbf{Z}^2 : \Lambda] = \mathrm{ord}(q)$. We aim to show that there are two linearly independent $(n_1, m_1), (n_2, m_2) \in \mathbf{Z}^2_{\geq 0} \cap \Lambda$ with $n_1, m_1, n_2, m_2 \leq C\,\mathrm{ord}(q)/\log q$ for some absolute constant C. Then we have

$$\gcd(2^{n_1}3^{m_1} - 1, 2^{n_2}3^{m_2} - 1) \geq q$$

and

$$\max(2^{n_1}3^{m_1} - 1, 2^{n_2}3^{m_2} - 1) \leq 3^{2C\,\mathrm{ord}(q)/\log q} \leq q^{2(\log 3)C\,\mathrm{ord}(q)/(\log q^2)}.$$

By Theorem 56 we must have then $\mathrm{ord}(q)/(\log q^2) \to 0$.

Let $(\widetilde{n}_1, \widetilde{m}_1), (\widetilde{n}_2, \widetilde{m}_2) \in \Lambda$ be two vectors that generate Λ and such that they have an angle as close to $\pi/2$ as possible. If the angle of $(\widetilde{n}_1, \widetilde{m}_1), (\widetilde{n}_2, \widetilde{m}_2)$ was less than $\pi/3$, then we could replace the longer vector by their difference and obtain a basis for Λ such that the angle of the two vectors are closer to $\pi/2$. This and a similar argument in the case where the angle is larger than $2\pi/3$ show that the angle must be between $\pi/3$ and $2\pi/3$. Since the area of the parallelepiped spanned by the two basis vectors is $\mathrm{ord}(q)$, we conclude

$$\|(\widetilde{n}_2, \widetilde{m}_2)\|_2 \cdot \|(\widetilde{n}_2, \widetilde{m}_2)\|_2 \leq \frac{\sqrt{3}}{2}\mathrm{ord}(q). \tag{29}$$

(This is a case of Minkowski's second theorem in the geometry of numbers.)

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

Combining this with (29), we get

$$\max(|\widetilde{n}_j|, |\widetilde{m}_j|) \le C_0 \frac{\mathrm{ord}(q)}{\log q}$$

for $j = 1, 2$ and some absolute constant C_0.

Now we have almost everything we want except that some of the n_j or m_j may be negative. To remedy this, suppose without loss of generality that $\|\widetilde{n}_1, \widetilde{m}_1\|_2 \ge \|\widetilde{n}_2, \widetilde{m}_2\|_2$, and consider the points $10(\widetilde{n}_1, \widetilde{m}_1) + k(\widetilde{n}_2, \widetilde{m}_2)$ for $k \in \mathbf{Z}$. A simple geometric argument shows that there are at least two among these whose coordinates are all of the same sign. We use these or their reflections to the origin as $(n_1, m_1), (n_2, m_2)$. $\square$

Now we turn to the proof of Theorem 56, 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 57. *Let $L \in \mathbf{Q}[X_1, \dots, X_d]$ be a linear form. Then there is a constant $C = C(L)$ such that*

$$L(x_1, \dots, x_d) = 0, \quad x_1, \dots, x_d \in \mathcal{S}$$

implies that

$$|x_i - x_j|_\infty |x_i - x_j|_2 |x_i - x_j|_3 \le C$$

for some $i \ne j$.

Remark 58. If $x \in \mathbf{Z}_{>0}$ and $x = 2^k 3^m y$ for some $k, m \in \mathbf{Z}_{\ge 0}$ and $2, 3 \nmid y$ then

$$|x|_\infty |x|_2 |x|_3 = y.$$

We restate the p-adic subspace theorem in the form we will use it.

Theorem 59. *Let V be a vector space of dimension n over $\overline{\mathbf{Q}}$. Let $S \subset M_{\mathbf{Q}}$ be a finite set of places, and for each $v \in S$, let $\Lambda_1^{(v)}, \dots \Lambda_n^{(v)} \in V^*$ be a basis. Furthermore, fix another basis $\Lambda_1^{(0)}, \dots, \Lambda_n^{(0)} \in V^*$. Fix an extension of $|\cdot|_v$ from $\mathbf{Q}$ to $\overline{\mathbf{Q}}$.*

Then for all $\varepsilon > 0$, there are finitely many $\varphi_1, \dots, \varphi_m \in V^*$ such that each solution $x \in V$ of the inequality

$$\prod_{v \in S} \prod_{j=1}^{n} |\Lambda_j(x)^{(v)}|_v \le H(\Lambda_1^{(0)}(x), \dots, \Lambda_n^{(0)}(x))^{-\varepsilon}, \qquad \Lambda_1^{(0)}(x), \dots, \Lambda_n^{(0)}(x) \in \mathbf{Z}$$

satisfies $\varphi_i(x) = 0$ for some $i = 1, \dots, m$.

Compared to Theorem 48 the role of V and V^* are interchanged. This formulation is a common generalization of 6 and 7. In fact, for our purposes in this section, working with a vector space over $\mathbf{Q}$ as in Theorem 7 would be enough, but we wanted to state include this general form somewhere in the course.

The construction of the linear forms in the following proof is due to Levin [8].

Proof of Theorem 56. Fix $\varepsilon > 0$. Let $a, b \in \mathcal{S}$ 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 ε.

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 independence of a and b.

We fix a suitably large $n \in \mathbf{Z}_{\geq 1}$. We will apply the subspace theorem on the vector space

$$V := \mathbf{Q}^{n^2}/\{(y, \ldots, y) : y \in \mathbf{Q}\}.$$

We will evaluate our functionals $\Lambda_j^{(v)}$ 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 $e_{n^2} = a^{n-1}b^{n-1}$.

We write $X_1, \ldots, X_{n^2}$ for the standard coordinates on $\mathcal{Q}^{n^2}$, and for the sake of concreteness, we write use $\Lambda_j^{(0)} = X_j - X_{n^2}$ for $j = 1, \ldots, n^2 - 1$ as our reference basis on V^*. Observe that these are well defined functionals on V. Note that $e_j \equiv 1 \mod d$ for all j, hence $e_i/d - e_j/d \in \mathbf{Z}$ for all i, j. Hence $\Lambda_j^{(0)}(e_1/d, \ldots, e_{n^2}/d) \in \mathbf{Z}$ for all j. We also record that

$$H(\Lambda_1^{(0)}(e_1/d, \ldots, e_{n^2}/d), \ldots, \Lambda_{n^2-1}^{(0)}(e_1/d, \ldots, e_{n^2}/d)) \leq \max(a,b)^{2n}.$$

We set $S = \{\infty, 2, 3\}$. To apply the subspace theorem, we select a basis $\Lambda_1^{(v)}, \ldots, \Lambda_{n^2-1}^{(v)}$ 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 our application of the subspace theorem, we choose each $\Lambda_j^{(v)}$ in the form $X_k - X_l$ for some k, l. We note the inequalities

$$|e_k/d - e_l/d|_v \leq \max(|e_k/d|_v, |e_l/d|_v). \tag{30}$$

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

Now we set $\Lambda_j^{(\infty)} = X_{j+1} - X_1$ for $j = 1, \ldots, n^2 - 1$. We note that $|e_1/d|_\infty$ is minimal among the $|e_j/d|_\infty$, hence (30) gives

$$|\Lambda_j^{(\infty)}(e_1/d, \ldots, e_{n^2}/d)|_\infty \leq |e_{j+1}/d|_\infty.$$

Therefore,

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

For $v = 2, 3$, we set $\Lambda_j^{(v)} = X_j - X_{n^2}$ for $j = 1, \ldots, n^2 - 1$. Using (30) and that $|e_j/d|_v$ is minimal for $j = n^2$, we get

$$\prod_{j=1}^{n^2-1} |\Lambda_j^{(v)}(e_1/d, \ldots, e_{n^2}/d)|_v \le |e_{n^2}/d|_v^{-1} \prod_{j=1}^{n^2} |e_j/d|_v = |a^{n-1}b^{n-1}|_v^{-1} \prod_{j=1}^{n^2} |e_j/d|_v.$$

In the last inequality we used $2, 3 \nmid d$.

Since each e_j is an integer divisible by no other primes than 2 and 3, we have $|e_j|_v = 1$ for all $v \notin S$. Therefore,

$$|e_j/d|_\infty |e_j/d|_2 |e_j/d|_3 = (\prod_{v \in M_{\mathbf{Q}}} |e_j|_v) |d^{-1}|_\infty = 1/d$$

for all j.

Using this and our previous bounds for the product corresponding to each place, we get

$$\prod_{v\in S} \prod_{j=1}^{n^2-1} |\Lambda_j^{(v)}(e_1/d, \ldots, e_{n^2}/d)|_v \le a^{n-1}b^{n-1}d \prod_{v\in S}\prod_{j=1}^{n^2} |e_j/d|_v = a^{n-1}b^{n-1}d^{-n^2+1}$$

Taking $n > 3\varepsilon^{-1} + 1$ and using $d > \max(a, b)^\varepsilon$, we get

$$a^{n-1}b^{n-1}d^{-n^2+1} \le a^{n-1}b^{n-1} \max(a, b)^{-3n} \le (\max(a, b)^{2n})^{-1/2}$$

and the subspace theorem applies with $1/2$ in the role of ε.

This means that there is a finite collection $\varphi_1, \ldots, \varphi_m \in V^*$ such that

$$\varphi_j(e_1/d, \ldots, e_{n^2}/d) = 0$$

for some j. These φ_j depend only on n and hence only on ε, and crucially not on a, b.

Now each φ_j induces a functional on $\mathbf{Q}^{n^2}$, which we denote by the same symbol, and we clearly have

$$\varphi_j(e_1, \ldots, e_{n^2}) = 0.$$

We apply Proposition 57 for each φ_j and find that there is a constant $C = C(\varepsilon)$ such that

$$|e_i - e_j|_\infty |e_i - e_j|_2 |e_i - e_j|_3 \le C$$

for some $i \ne j$. However, we have $d | e_i - e_j$, and also $e_i - e_j \ne 0$ by multiplicative independence of a, b, hence $d \le C$, as required. □

Proof of Proposition 57. 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 for x_1 and x_2. Therefore, we can find a suitably large C that works for all of these finitely many possibilities.

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 $x_1, \ldots, x_d \in \mathcal{S}$ 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 the greatest common divisor.

For simplicity, we assume $|x_j|_\infty$ is maximal for $j = d$. We let $w \in \{2, 3\}$ be such that $|x_d|_w \le |x_d|_\infty^{-1/2}$. 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.

We apply the subspace theorem on the vector space $V = \overline{\mathbf{Q}}^{d-1}$ with the standard basis $\Lambda_j^{(0)} = X_j$ as the reference basis. If we also take the linear forms $\Lambda_j^{(v)} = X_j$ for $v \in S = \{2, 3, \infty\}$ and $j = 1, \ldots, d-1$, then we get

$$\prod_{v \in S} \prod_{j=1}^{d-1} |\Lambda_j^{(v)}(x_1, \ldots, x_{d-1})|_v = 1.$$

Let $L = a_1 x_1 + \ldots + a_d x_d$. Recall the choice of $w = 2$ or 3 from above. If we replace $\Lambda_1^{(w)}$ by $(a_1/a_d)X_1 + \ldots + (a_{d-1}/a_d)X_{d-1}$, then we get

$$|\Lambda_1^{(w)}(x_1, \ldots, x_{d-1})|_w = |x_d|_w \le |x_d|_\infty^{-1/2}$$

instead of $|\Lambda_1^{(w)}(x_1, \ldots, x_{d-1})|_w = |x_1|_w = 1$. Therefore, for the modified $\Lambda_j^{(v)}$, we get

$$\prod_{v \in S} \prod_{j=1}^{d-1} |\Lambda_j^{(v)}(x_1, \ldots, x_{d-1})|_v \le |x_d|_\infty^{-1/2}.$$

We also note that

$$H(x_1, \ldots, x_{d-1}) \le |x_d|_\infty,$$

and hence the subspace theorem applies with $\varepsilon = 1/2$. 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. $\square$

7. 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 60 (Feldman). *Let α be an algebraic number of degree $d \geq 3$. Then there are effective constants $c = c(\alpha) > 0$ and $\varepsilon = \varepsilon(\alpha) > 0$ such that*

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

7.1. **Units.**

The purpose of this section is to prove the following auxiliary result, which is needed for the proof of Theorem 60.

Proposition 61. *For every number field K, there is an integer $r \in \mathbf{Z}_{\geq 0}$, $u_1 \dots, u_r \in K$ and an effective constant $C = C(K)$ such that the following holds. For every algebraic integer $\alpha \in \mathcal{O}_K$, there is another algebraic integer $\widetilde{\alpha} \in \mathcal{O}_K$ and integers $b_1, \dots, b_r \in \mathbf{Z}$ such that*

$$\begin{aligned} \alpha &= \widetilde{\alpha} u_1^{b_1} \cdots u_r^{b_r} \\ H(\widetilde{\alpha}) &\leq C |N_{K|\mathbf{Q}}(\alpha)|^{1/[K:\mathbf{Q}]} \\ |b_j| &\leq C \log H(\alpha). \end{aligned}$$

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

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

which is a homomorphism from the multiplicative group of K to the additive group $\mathbf{R}^{M_{K,\infty}}$.

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

It is immediate from the definitions that

$$|N_{K|\mathbf{Q}}(\alpha)| = \exp\Big(\sum_{v \in M_{K,\infty}} (\Phi(\alpha))_v \Big)$$

for all $\alpha \in K^\times$. In addition, for all $\alpha \neq 0 \in \mathcal{O}_K$, we have

$$H(\alpha)^d = \exp\Big(\sum_{v \in M_{K,\infty}} \max(0, (\Phi(\alpha))_v) \Big)$$

so

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

where $\|x\|_1 = \sum |x_v|$ on $\mathbf{R}^{M_{K,\infty}}$.

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 $|N_{K|\mathbf{Q}}(\alpha)| = 1$. We write W

for the 1-codimensional subspace of $\mathbf{R}^{M_{K,\infty}}$ satisfying the equation

$$\sum_v (x)_v = 0.$$

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

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

Proof of Proposition 61. Let $r = \dim W$ 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 $x \in \mathbf{R}^{M_{K,\infty}}$ be a vector with non-negative coordinates such that

$$\sum_{v \in M_{K,\infty}} x_v = \log N_{K|\mathbf{Q}}(\alpha).$$

Recall that $\alpha \in \mathcal{O}_K$, hence $N_{K|\mathbf{Q}}(\alpha) \geq 1$. There are many ways to choose x, and any choice will work.

Then $\Phi(\alpha) - x \in W$, so we can write

$$\Phi(\alpha) - x = y_1\Phi(u_1) + \ldots + y_r\Phi(u_r)$$

for some $y_j \in \mathbf{R}$. There is some constant C depending only on $u_1, \ldots, u_r$ so ultimately only on K such that

$$|y_j| \leq C\|\Phi(\alpha) - x\|_1 \leq 3Cd \log H(\alpha).$$

We define $b_j \in \mathbf{Z}$ with $|b_j| \leq |y_j|$ and $|b_j - y_j| \leq 1$, and set

$$\widetilde{\alpha} = \alpha u_1^{-b_1} \cdots u_r^{-b_r}.$$

Then $\widetilde{\alpha} \in \mathcal{O}_K$, and the required bound on $|b_j|$ holds.

Furthermore,

$$\begin{aligned} H(\widetilde{a})^d &\leq \exp(\|\Phi(\widetilde{\alpha})\|_1) = \exp(\|\sum_{j=1}^r (y_j - b_j)\Phi(u_j) + x\|_1) \\ &\leq \exp(C + \log N_{K|\mathbf{Q}}(\alpha)), \end{aligned}$$

where C is another constant depending only on K. Here we used that $\sum_{j=1}^r (y_j - b_j)\Phi(u_j)$ falls in a fixed bounded subset of W, so its norm can be bounded by a constant depending on K alone. □

7.2. **Proof of Theorem 60.** Before giving the proof, we discuss its strategy. We aim to find two algebraic numbers of the form

$$\alpha_1 \alpha_2^{b_2} \cdots \alpha_n^{b_n} \tag{31}$$

that are close to each other and then we will use an argument similar to the proof of Proposition 18. Thanks to Proposition 61, we know that algebraic integers of small norm are of this form.

Without loss of generality, we may assume that α is an algebraic integer. Let $P \in \mathbf{Z}[x]$ be the minimal polynomial of α. Write

$$P(x) = (x - \alpha_1) \cdots (x - \alpha_d)$$

with $\alpha_1 = \alpha$. Let $p/q \in \mathbf{Q}$ be such that it is closest to α_1 among the roots. As we have seen in Lemma 5, the absolute value of

$$(p - \alpha_1 q) \cdots (p - \alpha_d q) \tag{32}$$

is bounded above and below by a constant multiple of $q^d|\alpha_1 - p/q|$. Now we assume that (32) $\le q^\varepsilon$ 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 α.

Since $\alpha_1 \ldots, \alpha_d$ are algebraic integers, we have that $N_{\mathbf{Q}(\alpha)|\mathbf{Q}}(p-\alpha_j q)$ is an integer dividing $(32)^d$ for each j. This means that $p - \alpha_j q$ can be written in the form (31), 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 $|p - \alpha_1 q|$ is small (if $|\alpha_1 - p/q|$ is small) and hence $p - \alpha_j q$ is close to $(\alpha_1 - \alpha_j)q$ for each j. Therefore, $(\alpha_1-\alpha_2)(p-\alpha_3 q)$ and $(\alpha_1-\alpha_3)(p-\alpha_2 q)$ are close to each other because they are both close to $(\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)q$. On the other hand both $(\alpha_1 - \alpha_2)(p - \alpha_3 q)$ and $(\alpha_1 - \alpha_3)(p - \alpha_2 q)$ have small norm, so they can be written in the form (31).

We carry out the proof motivated by the above observations.

The proof uses estimates for linear forms in logarithms, which we recall now with heights of minimal polynomials replaced by Weil heights.

Theorem 62. *Let $n \in \mathbf{Z}_{\ge 1}$. Let $\alpha_1, \ldots, \alpha_n \in \overline{\mathbf{Q}}_{\neq 0}$ and let $\log \alpha_j$ be any choice of the logarithm of α_j. Let $b_1, \ldots, b_n \in \mathbf{Z}$ and let*

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

Let

$$A_j := \max(H(\alpha_j), \exp(|\log \alpha_j|), 10)$$

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

$$B^* := \max\Big(\frac{|b_1|}{\log A_n}, \ldots, \cdot\frac{|b_{n-1}|}{\log A_n}, |b_n|, 10\Big).$$

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

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

Proof of Theorem 60. We assume without loss of generality that α is an algebraic integer. We fix a small $\varepsilon > 0$. Let $p/q \in \mathbf{Q}$ be such that

$$|\alpha - p/q| < q^{-d+\varepsilon}.$$

Let

$$P(x) = (x - \alpha_1) \cdots (x - \alpha_d)$$

be the minimal polynomial of α, and let $\alpha_1 = \alpha$.

We compute

$$\begin{aligned}\Big|1-\frac{(\alpha_1-\alpha_3)(p-\alpha_2 q)}{(\alpha_1-\alpha_2)(p-\alpha_3 q)}\Big| =&\Big|1-\frac{(\alpha_1-\alpha_3)(p-\alpha_1 q+(\alpha_1-\alpha_2)q)}{(\alpha_1-\alpha_2)(p-\alpha_1 q+(\alpha_1-\alpha_3)q)}\Big|\\ \le&\Big|1-\frac{(\alpha_1-\alpha_3)(\alpha_1-\alpha_2)q}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)q}\Big|+Cq^{-d+\varepsilon}\\ =&Cq^{-d+\varepsilon}\end{aligned}$$

for a constant C depending only on α. In the rest of the proof C will denote a constant that depends only on α whose value may change between occurrences.

By Lemma 5, we have

$$|(p-\alpha_1 q)\cdots(p-\alpha_d q)|\le Cq^{\varepsilon}.$$

Since $p-\alpha_2 q$ is an algebraic integer, its norm (in the field extension $\mathbf{Q}(\alpha_2)|\mathbf{Q}$) is an integer that divides $(p-\alpha_1 q)^d\cdots(p-\alpha_d q)^d$. By Proposition 61, we can write

$$p-\alpha_2 q=\widetilde{\alpha}_2 u_1^{b_1}\cdots u_r^{b_r},$$

where $r\in\mathbf{Z}_{\ge 0}$, $u_1,\ldots,u_r$ are elements in $\mathbf{Q}(\alpha_2)$ independent of p and q, and $\widetilde{\alpha}_2\in\mathbf{Q}(\alpha_2)$, $b_1,\ldots,b_r\in\mathbf{Z}$ satisfy

$$H(\widetilde{\alpha}_2)\le Cq^{\varepsilon},\quad \max(|b_1|,\ldots,|b_r|)\le C\log q.$$

Here we used that $H(p-\alpha_2 q)\le(|p|+|q|)H(\alpha_2)\le Cq$.

Similarly, we can write

$$p-\alpha_2 q=\widetilde{\alpha}_3 w_1^{e_1}\cdots w_r^{e_r},$$

where $r\in\mathbf{Z}_{\ge 0}$, $w_1,\ldots,w_r$ are elements in $\mathbf{Q}(\alpha_3)$ independent of p and q, and $\widetilde{\alpha}_3\in\mathbf{Q}(\alpha_3)$, $e_1,\ldots,e_r\in\mathbf{Z}$ satisfy

$$H(\alpha_3)\le Cq^{\varepsilon},\quad \max(|e_1|,\ldots,|e_r|)\le C\log q.$$

Now we write

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

and note that

$$H(\alpha^*)\le Cq^{2\varepsilon}.$$

Plugging in to our estimate at the beginning of the proof, we get

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

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

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

where k is an integer satisfying

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

We apply Theorem 17 with α^* in the role of α_n and the u_j, w_j and -1 in the role of the other α_i. We have $A_i \le C$ for all $i \ne n$, $A_n \le Cq^{C\varepsilon}$ and

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

Theorem 62 yields

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

Choosing ε 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 α. This proves the theorem.

We used that the linear form in logarithms for which we applied Theorem 62 is non-zero. If it was 0, then we had

$$(\alpha_1 - \alpha_3)(p - \alpha_2 q) = (\alpha_1 - \alpha_2)(p - \alpha_3 q).$$

A simple calculation shows that this is equivalent to $\alpha_2 = \alpha_3$ or $p = \alpha_1 q$, and neither is the case. $\square$

If we apply the weaker bound in Theorem 16 instead of Theorem 62, we could get the slightly weaker conclusion

$$|\alpha - p/q| \ge cq^{-d + c/\log\log q}.$$

References

[1] A. Baker, *Transcendental number theory*, Second, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1990. MR1074572

[2] E. Bombieri and W. Gubler, *Heights in Diophantine geometry*, New Mathematical Monographs, vol. 4, Cambridge University Press, Cambridge, 2006. MR2216774

[3] Y. Bugeaud, *Linear forms in logarithms and applications*, IRMA Lectures in Mathematics and Theoretical Physics, vol. 28, European Mathematical Society (EMS), Zürich, 2018. MR3791777

[4] Y. Bugeaud, P. Corvaja, and U. Zannier, *An upper bound for the G.C.D. of* $a^n - 1$ *and* $b^n - 1$, Math. Z. **243** (2003), no. 1, 79–84. MR1953049

[5] J. W. S. Cassels, *An introduction to Diophantine approximation*, Hafner Publishing Co., New York, 1972. Facsimile reprint of the 1957 edition, Cambridge Tracts in Mathematics and Mathematical Physics, No. 45. MR0349591

[6] P. Corvaja and U. Zannier, *On the greatest prime factor of* $(ab + 1)(ac + 1)$, Proc. Amer. Math. Soc. **131** (2003), no. 6, 1705–1709. MR1955256

[7] S. Hernández and F. Luca, *On the largest prime factor of* $(ab+1)(ac+1)(bc+1)$, Bol. Soc. Mat. Mexicana (3) **9** (2003), no. 2, 235–244. MR2029272

[8] A. Levin, *Greatest common divisors and Vojta's conjecture for blowups of algebraic tori*, Invent. Math. **215** (2019), no. 2, 493–533. MR3910069

[9] D. Masser, *Auxiliary polynomials in number theory*, Cambridge Tracts in Mathematics, vol. 207, Cambridge University Press, Cambridge, 2016. MR3497545

[10] M. Nakamaye, *Roth's theorem: an introduction to diophantine approximation*, Rational points, rational curves, and entire holomorphic curves on projective varieties, 2015, pp. 75–108. MR3477541

[11] H. G. Senge and E. G. Straus, PV*-numbers and sets of multiplicity*, Period. Math. Hungar. **3** (1973), 93–100. MR340185

[12] C. L. Stewart, *On the representation of an integer in two different bases*, J. Reine Angew. Math. **319** (1980), 63–72. MR586115

[13] M. Waldschmidt, *Diophantine approximation on linear algebraic groups*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 326, Springer-Verlag, Berlin, 2000. Transcendence properties of the exponential function in several variables. MR1756786