Chapter 3

Roth's Theorem

3.1 Historical background (Liouville, Thue, Siegel, Gelfond, Dyson, Roth)

3.1.1 From Liouville to Thue

In Chapter 1, we proved the following Liouville's inequality on approximating algebraic numbers by rational numbers. The following statement is a reformulated version of Corollary 1.2.13.

Theorem 3.1.1 (Liouville). *Let $\alpha \in \mathbb{R}$ be an algebraic number of degree $d > 1$ over $\mathbb{Q}$. Then there exists a constant $c(\alpha) > 0$ such that for all rational numbers p/q $(q \geq 1)$, we have*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c(\alpha)}{q^d}. \tag{3.1.1}$$

In Chapter 1, we used the Fundamental Inequality (Proposition 1.2.10) to deduce this bound. In this chapter, we give another proof. This new proof sets up a prototype for various improvements on approximations of algebraic numbers by rational numbers, and will eventually lead to the deep Roth's Theorem and even more.

Proof. We will divide the proof into several steps.

Step I: Construct an auxiliary polynomial Let $f(x) \in \mathbb{Z}[x]$ be the minimal polynomial of α over $\mathbb{Q}$ with relatively prime integral coefficients. In particular, f is irreducible over $\mathbb{Q}$ and has degree d.

Step II: Non-vanishing at the rational point If $p/q \in \mathbb{Q}$, then we have $f(p/q) \neq 0$.

Step III: Lower bound (Liouville) By Step II, we then have $|f(p/q)| \geq 1/q^d$ since $\deg f = d$.

Step IV: Upper bound As $f(\alpha) = 0$ and f is the minimal polynomial of α, we can write $f(x) = (x - \alpha)g(x)$ with $g(\alpha) \neq 0$. Thus

$$\left|f\left(\frac{p}{q}\right)\right| = \left|\alpha - \frac{p}{q}\right| \cdot \left|g\left(\frac{p}{q}\right)\right|.$$

Notice that g has $d-1$ roots, and $\epsilon := \min_\beta |\beta - \alpha| > 0$ and $\delta := \max_\beta |\beta - \alpha| > 0$ with β running over all the roots of g. If $|p/q - \alpha| < \epsilon$, then $g(p/q) \neq 0$. Moreover, for any root β of g, we have $|p/q - \beta| \leq |\beta - \alpha| + |p/q - \alpha| \leq 2\delta$. Hence $0 \neq |g(p/q)| = \prod_\beta |p/q - \beta| \leq (2\delta)^{d-1}$ if $|p/q - \alpha| < \epsilon$. Notice that ϵ and δ are both determined by α.

Step V: Comparison of the two bounds The lower bound and the upper bound yield the following alternative: Either $|\alpha - p/q| \geq \epsilon \geq \epsilon/q^d$, or

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{1}{q^d}\frac{1}{(2\delta)^{d-1}}.$$

Thus it suffices to take $c(\alpha) = \min\{\epsilon, 1/(2\delta)^{d-1}\} > 0$. □

Before moving on, let us see an application. By this theorem of Liouville, one can see that $1 + \frac{1}{10^{2!}} + \frac{1}{10^{3!}} + \frac{1}{10^{4!}} + \cdots$ is a transcendental number since it has good rational approximations.

Improvements of Liouville's approximation above require sharpening the exponent on the right hand side of (3.1.1). The first improvement was obtained by Thue, replacing d by $\frac{d}{2}+1+\epsilon$.

Theorem 3.1.2 (Thue)**.** *Let $\alpha \in \mathbb{R}$ be an algebraic number of degree $d \geq 3$ over $\mathbb{Q}$ and let $\epsilon > 0$. Then there are only finitely many rational numbers p/q (with p, q coprime and $q \geq 1$) such that*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^{\frac{d}{2}+1+\epsilon}}. \tag{3.1.2}$$

Later on, Siegel improved this approximation by sharpening the exponent $\frac{d}{2} + 1 + \epsilon$ to $2\sqrt{d} + \epsilon$, which was further improved to $\sqrt{2d} + \epsilon$ by Gelfond and Dyson. The culminant of this approximation result is Roth's Theorem, replacing the exponent $\frac{d}{2} + 1 + \epsilon$ above by $2 + \epsilon$. Later on, a more general formulation of Roth's Theorem, concerning not only one but finitely many places, was obtained by Ridout over $\mathbb{Q}$ and by Lang over an arbitrary number field.

The proofs of these improvements follow the guideline set up above. In Liouville's work, the auxiliary polynomial from Step I comes for free and the polynomial has 1 variable. In general, we need to construct a polynomial such that the lower bound from Step III and the upper bound from Step IV repel each other.[1] This construction of the auxiliary polynomial is often by application of a suitable version of *Siegel's Lemma* discussed in Chapter 2. Thue and Siegel worked with polynomials in 2 variables. Roth obtained the drastic improvement by constructing a polynomial in m variables. However, the *non-vanishing of this auxiliary polynomial at a "special" point* from Step II is a crucial point of the construction and it is a major difficulty for the generalization of the approach. Solving this problem requires suitable *zero estimates* and even the more general *multiplicity estimates*, which themselves are an important topic of Diophantine Geometry.

Before moving on, let us see an example on how Thue's Theorem above can be applied to Diophantine equations. Stronger results on the finiteness of integer points on (certain) smooth affine curves can be obtained by applying Siegel's and Roth's Theorems.

Theorem 3.1.3. *Let $F(x, y) \in \mathbb{Z}[x, y]$ be a homogeneous polynomial of degree d with at least 3 non-proportional linear factors over $\mathbb{C}$. Then for every non-zero $m \in \mathbb{Z}$, the equation $F(x, y) = m$ has only finitely many integer solutions.*

Proof. We prove this by contradiction. First assume that F is irreducible over $\mathbb{Q}$. Consider the decomposition over $\mathbb{C}$

$$F\left(\frac{x}{y}, 1\right) = a_d\left(\frac{x}{y} - \alpha_1\right)\cdots\left(\frac{x}{y} - \alpha_d\right).$$

[1] We will see more precise meaning of this in later sections; a notion of "index" will be used.

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

$$a_d\left(\frac{x}{y}-\alpha_1\right)\cdots\left(\frac{x}{y}-\alpha_d\right)=\frac{m}{y^d}.$$

If it has infinitely many integer solutions (x_n,y_n), then $|y_n|\to\infty$ and hence $m/y_n^d\to 0$. Thus up to passing to a subsequence, we may and do assume that $x_n/y_n\to\alpha_j$ for some j. Notice that $|x_n/y_n-\alpha_i|>\epsilon$ for some ϵ depending only on F for all $i\neq j$. Thus we obtain infinitely many integral solutions to $|\alpha_j-p/q|\leq Cq^{-d}$ for some constant $C>0$. This contradicts Thue's Theorem above since $d\geq 3$.

Next we pass to the general case. Let $F_1,\ldots,F_r$ be the distinct non-constant irreducible polynomials in $\mathbb{Z}[x,y]$ dividing F. By a linear change of coordinates, we may and do assume that the polynomial y does not divide F. Assume $F(x,y)=m$ has infinitely many integer solutions. By the Pigeonhole Principle, there exist divisors $m_1,\ldots,m_r$ of m with the following property: the system $F_1(x,y)=m_1,\ldots,F_r(x,y)=m_r$ has infinitely integer solutions (x_n,y_n). As in the previous case, up to passing to a subsequence we may and do assume that x_n/y_n converges to a root of $F_j(x,1)$ for each $j\in\{1,\ldots,r\}$. But the F_j's have distinct roots since each F_j is the minimal polynomial of each one of its roots. So $r=1$. By the assumption that F has at least 3 non-proportional linear factors over $\mathbb{C}$, we then have $\deg F_1\geq 3$. Thus the conclusion follows from the irreducible case applied to $F_1(x,y)=m_1$. □

3.1.2 Statement of Roth's Theorem

The original version of Roth's Theorem, which we will prove in this chapter, is as follows.

Theorem 3.1.4 (Roth's Theorem)**.** *Let $\alpha\in\mathbb{R}$ be an algebraic number and let $\epsilon>0$. Then there are only finitely many rational numbers p/q (with p,q coprime and $q\geq 1$) such that*

$$\left|\alpha-\frac{p}{q}\right|\leq\frac{1}{q^{2+\epsilon}}. \tag{3.1.3}$$

A more general version by Lang is as follows. The statement uses the multiplicative height H.

Theorem 3.1.5. *Let K be a number field and let $S\subseteq M_K$ a finite subset. For each $v\in S$, take $\alpha_v\in K_v$ which is K-algebraic, i.e. $\alpha_v\in K_v$ is a root of a polynomial with coefficients in K. Then for each $\epsilon>0$, there are only finitely many $\beta\in K$ such that*

$$\prod_{v\in S}\min\{1,|\alpha_v-\beta|_v\}\leq H(\beta)^{-(2+\epsilon)}. \tag{3.1.4}$$

Implication of Theorem 3.1.4 by Theorem 3.1.5. Take $K=\mathbb{Q}$ and $S=\{\infty\}$. Then (3.1.4) implies that there are only finitely many rational numbers p/q such that $\min\{1,|\alpha-p/q|\}\leq H(p/q)^{-(2+\epsilon)}$. Recall that $H(p/q)\geq 1$. So if $\min\{1,|\alpha-p/q|\}\leq H(p/q)^{-(2+\epsilon)}$, then $|\alpha-p/q|\leq 1$. Therefore, there are only finitely many rational numbers p/q (with p,q coprime and $q\geq 1$) such that $|\alpha-p/q|\leq\max\{|p|,q\}^{-(2+\epsilon)}=\min\{|p|^{-(2+\epsilon)},q^{-(2+\epsilon)}\}\leq q^{-(2+\epsilon)}$. This proves Theorem 3.1.4. □

3.2 Index and preparation of the construction of the auxiliary polynomial

In the Thue–Siegel method and Roth's proof of his big theorem, it is important to construct a polynomial of *rapid decreasing degrees*, for the purpose of making the lower bound and the upper bound repel each other. Then, in order to say that the polynomial vanishes at high order, we need a suitable notion of *index*.

Let F be a field. Let $P \in F[x_1, \dots, x_m]$ be a polynomial in m variables. Let $\mathbf{d} = (d_1, \dots, d_m)$ be an m-uple (warning: the d_j's may not be the partial degrees of P). Denote by $\mathbf{x} = (x_1, \dots, x_m)$.

To ease notation, we introduce the following abbreviation. For two m-uples $\mathbf{n} = (n_1, \dots, n_m)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$ of non-negative integers, set

$$\binom{\mathbf{n}}{\boldsymbol{\mu}} = \prod_{j=1}^{m} \binom{n_j}{\mu_j}$$

and

$$\partial_{\boldsymbol{\mu}} = \frac{1}{\mu_1! \cdots \mu_m!} \left(\frac{\partial}{\partial x_1}\right)^{\mu_1} \cdots \left(\frac{\partial}{\partial x_m}\right)^{\mu_m}.$$

Then

$$\partial_{\boldsymbol{\mu}} \mathbf{x}^{\mathbf{n}} = \binom{\mathbf{n}}{\boldsymbol{\mu}} \mathbf{x}^{\mathbf{n}-\boldsymbol{\mu}}.$$

The following lemma is useful. It will be proved in the Exercise class.

Lemma 3.2.1. *$h(\partial_{\boldsymbol{\mu}} P) \leq h(P) + (\deg P) \log 2$ where $\deg P$ is the sum the partial degrees of P.*

Now let us define the index.

Definition 3.2.2. *For a point $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$, the **index** of P at $\boldsymbol{\alpha}$ with respect to $\mathbf{d}$ is defined to be*

$$\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) := \min_{\boldsymbol{\mu}} \left\{ \frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m} : \partial_{\boldsymbol{\mu}} P(\boldsymbol{\alpha}) \neq 0 \right\}. \tag{3.2.1}$$

Another way to see the index is by writing P to be $P = \sum_{\boldsymbol{\mu}} b_{\boldsymbol{\mu}} (x_1 - \alpha_1)^{\mu_1} \cdots (x_m - \alpha_m)^{\mu_m}$, and then $\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) = \min\{\sum_{j=1}^{m} \frac{\mu_j}{d_j} : b_{\boldsymbol{\mu}} \neq 0\}$.

Lemma 3.2.3. *The following properties hold true.*

(i) $\operatorname{ind}(P + Q; \mathbf{d}; \boldsymbol{\alpha}) \geq \min\{\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}), \operatorname{ind}(Q; \mathbf{d}; \boldsymbol{\alpha})\}$;

(ii) $\operatorname{ind}(PQ; \mathbf{d}; \boldsymbol{\alpha}) = \operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) + \operatorname{ind}(Q; \mathbf{d}; \boldsymbol{\alpha})$;

(iii) $\operatorname{ind}(\partial_{\boldsymbol{\mu}} P; \mathbf{d}; \boldsymbol{\alpha}) \geq \operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) - \frac{\mu_1}{d_1} - \cdots - \frac{\mu_m}{d_m}$.

Proof. For (i): Assume that $\operatorname{ind}(P+Q; \mathbf{d}; \boldsymbol{\alpha})$ is achieved at some $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$, then $\partial_{\boldsymbol{\mu}}(P+Q)(\boldsymbol{\alpha}) \neq 0$. So $\partial_{\boldsymbol{\mu}} P(\boldsymbol{\alpha}) + \partial_{\boldsymbol{\mu}} Q(\boldsymbol{\alpha}) \neq 0$, and therefore either $\partial_{\boldsymbol{\mu}} P(\boldsymbol{\alpha}) \neq 0$ or $\partial_{\boldsymbol{\mu}} Q(\boldsymbol{\alpha}) \neq 0$. By definition of the index, we then have: either $\sum \frac{\mu_j}{d_j} \geq \operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha})$ or $\sum \frac{\mu_j}{d_j} \geq \operatorname{ind}(Q; \mathbf{d}; \boldsymbol{\alpha})$. Thus $\operatorname{ind}(P + Q; \mathbf{d}; \boldsymbol{\alpha}) = \sum \frac{\mu_j}{d_j} \geq \min\{\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}), \operatorname{ind}(Q; \mathbf{d}; \boldsymbol{\alpha})\}$.

For (ii): Assume that $\operatorname{ind}(PQ; \mathbf{d}; \boldsymbol{\alpha})$ is achieved at some $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$. We have $\partial_{\mu}(PQ) = \sum_{\boldsymbol{\mu_1} + \boldsymbol{\mu_2} = \boldsymbol{\mu}} C_{\boldsymbol{\mu_1}, \boldsymbol{\mu_2}} (\partial_{\boldsymbol{\mu_1}} P)(\partial_{\boldsymbol{\mu_2}} Q)$ for some positive integers $C_{\boldsymbol{\mu_1}, \boldsymbol{\mu_2}}$.[2] Thus there

[2] In fact, it can be checked that each $C_{\boldsymbol{\mu_1}, \boldsymbol{\mu_2}}$ is equal to 1.

exists $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ such that $\boldsymbol{\mu_1}+\boldsymbol{\mu_2}=\boldsymbol{\mu}$, $\partial_{\boldsymbol{\mu_1}}P(\boldsymbol{\alpha})\neq 0$ and $\partial_{\boldsymbol{\mu_2}}Q(\boldsymbol{\alpha})\neq 0$. Thus the definition of index yields $\sum_j \frac{\mu_{1,j}}{d_j}\geq \mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})$ and $\sum_j \frac{\mu_{2,j}}{d_j}\geq \mathrm{ind}(Q;\mathbf{d};\boldsymbol{\alpha})$. So $\mathrm{ind}(PQ;\mathbf{d};\boldsymbol{\alpha})=\sum_j \frac{\mu_{1,j}+\mu_{2,j}}{d_j}\geq \mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})+\mathrm{ind}(Q;\mathbf{d};\boldsymbol{\alpha})$.

To get the other direction, let us look at the set of $\boldsymbol{\mu_1}$'s such that

$$\partial_{\boldsymbol{\mu_1}}P(\boldsymbol{\alpha})\neq 0 \quad \text{and} \quad \mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})=\sum_j \frac{\mu_{1,j}}{d_j}.$$

Consider the *smallest* such m-uple, ordered by the lexicographic order, which we call $\boldsymbol{\nu_1}$. Similarly take $\boldsymbol{\nu_2}$ for Q. Set $\boldsymbol{\nu}=\boldsymbol{\nu_1}+\boldsymbol{\nu_2}$. Then

$$\partial_{\boldsymbol{\nu}}(PQ)(\boldsymbol{\alpha})=C_{\boldsymbol{\nu_1},\boldsymbol{\nu_2}}\partial_{\boldsymbol{\nu_1}}P(\boldsymbol{\alpha})\cdot\partial_{\boldsymbol{\nu_2}}Q(\boldsymbol{\alpha})$$

because all the other terms vanish! Thus $\mathrm{ind}(PQ;\mathbf{d};\boldsymbol{\alpha})\leq\sum_j\frac{\nu_j}{d_j}=\sum_j\frac{\nu_{1,j}+\nu_{2,j}}{d_j}=\mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})+\mathrm{ind}(Q;\mathbf{d};\boldsymbol{\alpha})$. Hence we are done by the previous paragraph.

For (iii): Assume that $\mathrm{ind}(\partial_{\boldsymbol{\mu}}P;\mathbf{d};\boldsymbol{\alpha})$ is achieved at some $\boldsymbol{\nu}=(\nu_1,\dots,\nu_m)$. Then $\partial_{\boldsymbol{\nu}}(\partial_{\boldsymbol{\mu}}P)(\boldsymbol{\alpha})\neq 0$, and hence $\partial_{\boldsymbol{\nu}+\boldsymbol{\mu}}P(\boldsymbol{\alpha})\neq 0$. So $\sum_j\frac{\nu_j+\mu_j}{d_j}\geq\mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})$. Hence $\mathrm{ind}(\partial_{\boldsymbol{\mu}}P;\mathbf{d};\boldsymbol{\alpha})=\sum_j\frac{\nu_j}{d_j}\geq\mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})-\sum_j\frac{\mu_j}{d_j}$. □

Our purpose is to find a polynomial of large index and of small height. The result is as follows. Set, for each $t>0$,

$$\mathcal{V}_m(t):=\{\mathbf{x}\in\mathbb{R}^m : x_1+\cdots+x_m\leq t,\ 0\leq x_j\leq 1\},$$

and $V_m(t)$ to be the volume of $\mathcal{V}_m(t)$ with respect to the usual Lebesgue measure on $\mathbb{R}^m$.

Lemma 3.2.4. *Let $\alpha\in\mathbb{R}$ be an algebraic number, and set $\boldsymbol{\alpha}=(\alpha,\dots,\alpha)\in\mathbb{R}^m$. Let $r=[\mathbb{Q}(\alpha):\mathbb{Q}]$. Let $t>0$ be such that $rV_m(t)<1$. Then, for all sufficiently large integers $d_1,\dots,d_m$, there exists a polynomial $P\in\mathbb{Q}[x_1,\dots,x_m]$ of partial degrees at most $d_1,\dots,d_m$ such that:*

(i) $\mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})\geq t$;

(ii) as $d_j\to\infty$ for all $j\in\{1,\dots,m\}$, we have

$$h(P)\leq\frac{rV_m(t)}{1-rV_m(t)}\sum_{j=1}^m(h(\alpha)+\log 2+o(1))d_j.$$

Proof. The key ingredient to prove this lemma is by applying Siegel's Lemma (and it suffices to apply the basic relative version, Lemma 2.1.3). Let us explain what the parameters and the linear system from Siegel's Lemma are in the current situation.

Write $P(\mathbf{x})=\sum p_J\mathbf{x}^J$ for the polynomial. Then any P with $\mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha})\geq t$ lies in the set of P satisfying

$$\partial_I P(\boldsymbol{\alpha})=0 \quad \text{for all} \quad \frac{i_1}{d_1}+\cdots+\frac{i_m}{d_m}<t \tag{3.2.2}$$

with $I=(i_1,\dots,i_m)$. Notice that we may assume $i_k\leq d_k$ for each $k\in\{1,\dots m\}$ because otherwise the partial derivative will be identically 0. Now all the equations from (3.2.2) define a linear system A in the coefficients p_J of P which we wish to solve in $\mathbb{Q}$.

Each entry in this linear system A is of the form $\binom{J}{I}\boldsymbol{\alpha}^{J-I}$, and thus $H(A)\leq 2^{d_1+\cdots+d_m}H(\alpha)^{d_1+\cdots+d_m}$.

The number N of unknowns is $N=(d_1+1)\cdots(d_m+1)$. Notice that $N\sim d_1\cdots d_m$ as $d_j\to\infty$ for all $j\in\{1,\dots,m\}$.

The number M of equations is $M = \#(\Gamma \cap \mathcal{V}_m(t))$ for the lattice $\Gamma = \frac{1}{d_1}\mathbb{Z} \times \cdots \frac{1}{d_m}\mathbb{Z}$. We claim that $M \sim V_m(t)d_1 \cdots d_m$ as $d_j \to \infty$ for all $j \in \{1, \ldots, m\}$. Indeed, $V_m(t)d_1 \cdots d_m \leq M$ because we can associate to each lattice point in Γ the parallelopiped $[i_1/d_1, (i_1+1)/d_1] \times \cdots \times [i_m/d_m, (i_m+1)/d_m]$. On the other hand, for each $(i_1/d_1, \ldots, i_m/d_m) \in \Gamma \cap \mathcal{V}_m(t)$, we have

$$\frac{i_1+1}{d_1} + \cdots + \frac{i_m+1}{d_m} \leq t + \frac{1}{d_1} + \cdots + \frac{1}{d_m} \quad \text{and} \quad i_j + 1 \leq d_j + 1.$$

Thus if we rescale $\mathcal{V}_m(t)$ by the factor $1 + \max\{1, t^{-1}\}(1/d_1 + \cdots + 1/d_m)$, then the rescaled domain contains all the parallelopipeds associated to the points in $\Gamma \cap \mathcal{V}_m(t)$. In summary, we have

$$V_m(t)d_1 \cdots d_m \leq M \leq V_m(t)\left(1 + \max\{1, t^{-1}\}\left(\frac{1}{d_1} + \cdots + \frac{1}{d_m}\right)\right)^m d_1 \cdots d_m.$$

Thus $M \sim V_m(t)d_1 \cdots d_m$ as $d_j \to \infty$ for all $j \in \{1, \ldots, m\}$.

Now we are ready to apply Siegel's Lemma. As $d_j \to \infty$ for all $j \in \{1, \ldots, m\}$, we have $N \sim d_1 \cdots d_m > rM$ because $rV_m(t) < 1$. Thus by Lemma 2.1.3 and the comment below (which relates the right hand side of the height bound to the height of the matrix by using the Fundamental Inequality Proposition 1.2.10), there is a non-zero solution to the linear system defined by (3.2.2), and hence a non-zero polynomial P satisfying hypothesis (i), such that (for some constant C depending only on α)

$$\begin{aligned} h(P) &\leq \frac{rV_m(t)d_1 \cdots d_m}{d_1 \cdots d_m - rV_m(t)d_1 \cdots d_m} \log(Cd_1 \cdots d_m H(A)) \\ &\leq \frac{rV_m(t)}{1 - rV_m(t)} \left(\sum_{j=1}^m \log d_j + (h(\alpha) + \log 2)\sum_{j=1}^m d_j + \log C\right) \end{aligned}$$

as $d_j \to \infty$ for all $j \in \{1, \ldots, m\}$. Hence we are done. □

Next we give an estimate of the volume in question.

Lemma 3.2.5. *If $0 \leq \epsilon \leq 1/2$, then*

$$V_m\left(\left(\frac{1}{2} - \epsilon\right)m\right) \leq e^{-6m\epsilon^2}.$$

Proof. Set $\chi(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x \geq 0 \end{cases}$. Then $\chi(x) < e^{-\lambda x}$ for every $\lambda > 0$. Thus for each $\lambda > 0$, we have

$$\begin{aligned} V_m\left(\left(\frac{1}{2} - \epsilon\right)m\right) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \cdots \int_{-\frac{1}{2}}^{\frac{1}{2}} \chi(x_1 + \cdots + x_m + m\epsilon)\mathrm{d}x_1 \cdots \mathrm{d}x_m \\ &\leq \int_{-\frac{1}{2}}^{\frac{1}{2}} \cdots \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-\lambda(m\epsilon + \sum x_j)}\mathrm{d}x_1 \cdots \mathrm{d}x_m \\ &= \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-\lambda(\epsilon + x)}\mathrm{d}x\right)^m \\ &= e^{-mU(\lambda)}, \end{aligned}$$

where $U(\lambda) = \epsilon\lambda - \log\frac{\sinh(\lambda/2)}{\lambda/2}$.[3] But $\sinh(u)/u = 1 + u^2/3! + u^4/5! + \cdots \leq 1 + u^2/6 + (u^2/6)^2/2! + \cdots = e^{u^2/6}$. So we can conclude by setting $\lambda = 12\epsilon$. □

[3] $\sinh(u) = \frac{e^u - e^{-u}}{2}$.

3.2.1 Why does it help to have more variables in the construction of auxiliary polynomial?

Let $\alpha \in \mathbb{R}$ be an algebraic number of degree d. We wish to show that α cannot be well approximated by rational numbers. In more vigorous terms, this means that we wish to obtain a result of the following type: There are only finitely many rational numbers p/q such that $|\alpha - \frac{p}{q}| \leq \frac{1}{q^\kappa}$; here κ is a constant and we wish to give the best possible κ. Now let us see how the *Law of Large Numbers* yields the following philosophy: For the construction of the auxiliary polynomial $P \in \mathbb{Z}[x_1, \ldots, x_m]$ (for example as in Lemma 3.2.4), when $m \to \infty$ the exponent κ becomes better. And in the end, we see why $2+\epsilon$ is the best possible exponent in this theoretical way. This is via the *index*.

Assume we find $P \in \mathbb{Z}[x_1, \ldots, x_n]$ such that $P(\alpha, \cdots, \alpha) = 0$ and $P(p_1/q_1, \ldots, p_m/q_m) \neq 0$ with $|\alpha - p_i/q_i| < 1/q_i^\kappa$. Assume P has partial degrees at most $\mathbf{d} = (d_1, \ldots, d_m)$.

Let us study the index $\mathrm{ind}(P, \mathbf{d}; \boldsymbol{\alpha})$ where $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)$, using the comment below (3.2.1). A monomial in $x_1 - \alpha, \ldots, x_m - \alpha$ is an m-tuple $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ with $0 \leq \mu_j \leq d_j$. For each j, roughly half of the possible μ_js satisfy $\frac{\mu_j}{d_j} \geq \frac{1}{2}$ and the other half satisfy $\frac{\mu_j}{d_j} < \frac{1}{2}$. Moreover, the possible values of $\frac{\mu_j}{d_j}$ are evenly distributed in $[0, 1]$. Therefore, for a randomly chosen $\boldsymbol{\mu}$, the expected value of $\sum \frac{\mu_j}{d_j}$ is $\frac{m}{2}$. So the (weak) Law of Large Numbers yields: For each $\epsilon' > 0$, we have

$$\frac{\#\{\boldsymbol{\mu} : \sum \frac{\mu_j}{d_j} < \frac{m}{2}(1-\epsilon')\}}{(d_1+1)\cdots(d_m+1)} \to 0 \quad \text{as } m \to \infty. \tag{3.2.3}$$

Thus when m gets larger and larger, there are more and more polynomials of partial degrees at most $(d_1, \ldots, d_m)$ whose index at $(\alpha, \ldots, \alpha)$ is $\geq \frac{m}{2}(1-\epsilon')$. This also explains the parameter chosen for the volume estimate in Lemma 3.2.5. In later sections, we will see that the desired κ is expected to be $m/\frac{m}{2}(1-\epsilon')$, which is then of the form $2+\epsilon$ for some $\epsilon > 0$.

3.3 Proof of Roth's Theorem assuming zero estimates

In this section, we prove Roth's Theorem (Theorem 3.1.4) *assuming zero estimates*. The result for zero estimates which we will cite is *Roth's Lemma*.

We start by restating Roth's Theorem.

Theorem 3.3.1 (Roth's Theorem)**.** *Let $\alpha \in \mathbb{R}$ be an algebraic number and let $\epsilon > 0$. Then there are only finitely many rational numbers p/q (with p, q coprime and $q \geq 1$) such that*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^{2+\epsilon}}. \tag{3.3.1}$$

We will divide the proof into several step, outlined as for Theorem 3.1.1.

3.3.0 Step 0: Choosing independent solutions.

Assume the conclusion is wrong. Then there exists $\alpha \in \mathbb{R}$ an algebraic number with infinitely many rational approximations p/q to α satisfying (3.3.1). Then, for any positive integer m and any large constants L and M, we can find m such rational approximations p_j/q_j to α (with $q_j \geq 1$) such that

$$\log q_1 > L \quad \text{and} \quad \log q_{j+1} > M \log q_j$$

for each $j \in \{1, \ldots, m-1\}$. Namely, we consider *large solutions* which satisfy a *Gap Principle*.

Such a sequence will be called **$(\mathbf{L}, \mathbf{M})$-independent**.

Fix $\epsilon' \in (0, 1/6)$.

3.3.1 Step 1: Construction of an auxiliary polynomial.

Let D be a large real number which we will fix later on. For each $j \in \{1, \dots, m\}$, set

$$d_j := \lfloor D/\log q_j \rfloor.$$

In this step, we wish to construct a polynomial $P(\mathbf{x}) \in \mathbb{Z}[\mathbf{x}] = \mathbb{Z}[x_1, \dots, x_m]$ of partial degrees $d_1, \dots, d_m$, vanishing to a (weighted) high order at $\boldsymbol{\alpha} = (\alpha, \dots, \alpha)$. More precisely, we will apply Lemma 3.2.4[4] to construct a polynomial P of large index at $\boldsymbol{\alpha}$ with respect to $\mathbf{d}$. More precisely, Lemma 3.2.5 implies $V_m((1/2 - \epsilon')m) \le e^{-6m\epsilon'^2}$. If we choose

$$m > \frac{\log 2[\mathbb{Q}(\alpha) : \mathbb{Q}]}{6\epsilon'^2}, \tag{3.3.2}$$

then $[\mathbb{Q}(\alpha) : \mathbb{Q}]V_m(t) \le 1/2$. Thus Lemma 3.2.4 yields a polynomial P of partial degrees at most $d_1, \dots, d_m$ such that:

(i) $\mathrm{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) \ge (1/2 - \epsilon')m$, or equivalently for any $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$ with

$$\frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m} < \left(\frac{1}{2} - \epsilon'\right) m$$

satisfies $\partial_{\boldsymbol{\mu}} P(\boldsymbol{\alpha}) = 0$;

(ii) As $d_j \to \infty$ for all $j \in \{1, \dots, m\}$, we have

$$h(P) \le \sum_{j=1}^{m} (h(\alpha) + \log 2 + o(1)) d_j \le C(d_1 + \cdots + d_m) \tag{3.3.3}$$

with C a suitable constant depending only on α and m.

3.3.2 Step 2: Non-vanishing at the rational points.

This is the most difficult step. Before Roth's work, one could only do for $m = 1$ and $m = 2$. Roth proved, for this step, the following lemma as a consequence of *Roth's Lemma*. It is in this step that we need the parameter M; see (3.4.2). Notice also that all the conditions for the parameters (m, L, M and D) are summarized in the hypotheses of this lemma.

Lemma 3.3.2. *Suppose $p_1/q_1, \dots, p_m/q_m$ are (L, M)-independent with*

$$m > \log(2[\mathbb{Q}(\alpha) : \mathbb{Q}])/(6\epsilon'^2) \quad \text{and} \quad L \ge (C+4)m\epsilon'^{-2^{m-1}} \quad \text{and} \quad M \ge 2\epsilon'^{-2^{m-1}}.$$

Then for every sufficiently large D, there exists a polynomial $Q \in \mathbb{Z}[x_1, \dots, x_m]$ with partial degrees at most $d_j = \lfloor D/\log q_j \rfloor$ such that

(i) $\mathrm{ind}(Q; \mathbf{d}; \boldsymbol{\alpha}) \ge \left(\frac{1}{2} - 3\epsilon'\right) m$;

(ii) $Q(p_1/q_1, \dots, p_m/q_m) \ne 0$;

(iii) $h(Q) \le C_1 m D/L$ *for a constant C_1 depending only on α and m.*

In fact, this Q is a suitable derivative of the P constructed from Step 1.

[4] Which itself is a suitable application of Siegel's Lemma.

3.3.3 Step 3: Lower bound (Liouville).

Since $Q(p_1/q_1,\dots,p_m/q_m)\neq 0$ and Q has partial degrees at most $d_1,\dots,d_m$, we have the obvious bound (Liouville bound)

$$\log|Q(p_1/q_1,\dots,p_m/q_m)|\geq \log q_1^{-d_1}\cdots q_m^{-d_m}=-(d_1\log q_1+\ldots+d_m\log q_m).$$

The choice $d_j=\lfloor D/\log q_j\rfloor$ implies $D-\log q_j\leq d_j\log q_j\leq D$. Thus

$$\log|Q(p_1/q_1,\dots,p_m/q_m)|\geq -mD.$$

3.3.4 Step 4: Upper bound.

Consider the Taylor expansion of Q at $(\alpha,\dots,\alpha)$. Since $\mathrm{ind}(Q;\mathbf{d};\boldsymbol{\alpha})\geq\left(\frac{1}{2}-3\epsilon'\right)m$, we get

$$Q(p_1/q_1,\dots,p_m/q_m)=\sum \partial_{\boldsymbol{\mu}}Q(\boldsymbol{\alpha})(p_1/q_1-\alpha)^{\mu_1}\cdots(p_m/q_m-\alpha)^{\mu_m} \tag{3.3.4}$$

with $\boldsymbol{\mu}=(\mu_1,\dots,\mu_m)$ running over all possibilities with $\sum_j \mu_j/d_j\geq(1/2-3\epsilon')m$. Then the assumption $|\alpha-p_j/q_j|\leq q_j^{-(2+\epsilon)}$ implies

$$\begin{aligned}\log\left(|p_1/q_1-\alpha|^{\mu_1}\cdots|p_m/q_m-\alpha|^{\mu_m}\right)&\leq\sum_j\frac{\mu_j}{d_j}\log q_j^{-(2+\epsilon)d_j}\\&\leq(\max_j\log q_j^{-(2+\epsilon)d_j})\sum_j\frac{\mu_j}{d_j}\\&\leq(2+\epsilon)(1/2-3\epsilon')m\max_j\{-d_j\log q_j\}\\&=-(2+\epsilon)(1/2-3\epsilon')m\min_j d_j\log q_j\\&\leq-(2+\epsilon)(1/2-3\epsilon')m(D-\log q_m).\end{aligned}$$

Now let us estimate $\log|\partial_{\boldsymbol{\mu}}Q(\boldsymbol{\alpha})|$. We use Lemma 3.2.1 and Proposition 1.3.2 to get

$$\begin{aligned}h(\partial_{\boldsymbol{\mu}}Q(\boldsymbol{\alpha}))&\leq h(Q)+(\log 2)\sum_j d_j+h(\alpha)\sum_j d_j+(m+\sum_j d_j+1)\log 2\\&\leq C_1\frac{mD}{L}+(h(\alpha)+\log 4)\sum d_j+(m+1)\log 2.\end{aligned}$$

The Fundamental Inequality, Proposition 1.2.10, yields $\log|\partial_{\boldsymbol{\mu}}Q(\boldsymbol{\alpha})|\leq h(\partial_{\boldsymbol{\mu}}Q(\boldsymbol{\alpha}))$. As $d_j=\lfloor D/\log q_j\rfloor\leq D/\log q_j\leq D/\log q_1<D/L$ (recall that $\log q_j\geq\log q_1>L$), we have

$$\log|\partial_{\boldsymbol{\mu}}Q(\boldsymbol{\alpha})|\leq(C_1+h(\alpha)+\log 4)\frac{mD}{L}+(m+1)\log 2.$$

Notice that the number of terms in the expression of $Q(p_1/q_1,\dots,p_m/q_m)$ from (3.3.4) is polynomial in $d_1,\dots,d_m$, and hence the contribution of this number to $\log|Q(p_1/q_1,\dots,p_m/q_m)|$ is $o(d_1+\cdots+d_m)=o(mD/L)$. Thus

$$\log|Q(p_1/q_1,\dots,p_m/q_m)|\leq C'\frac{mD}{L}+(m+1)\log 2-(2+\epsilon)(\frac{1}{2}-3\epsilon')m(D-\log q_m)$$

for a suitable constant C' depending only on α and m.

3.3.5 Step 5: Comparison of the two bounds.

Now the two bounds from Step 3 and Step 4 together imply

$$mD \geq (2+\epsilon)\left(\frac{1}{2}-3\epsilon'\right)m(D-\log q_m) - C'\frac{mD}{L} - (m+1)\log 2.$$

Dividing both sides by mD, we get

$$1 \geq (2+\epsilon)\left(\frac{1}{2}-3\epsilon'\right)\left(1-\frac{\log q_m}{D}\right) - \frac{C'}{L} - \frac{(m+1)\log 2}{mD}.$$

Recall that q_m is fixed. Now let $\epsilon' \to 0$, $D \to \infty$ and $L \to \infty$. Then we get $1 \geq 1+\epsilon/2$. This is a contradiction. Hence we are done. □

Remark 3.3.3. *In this proof, we gave an explicit bound for $d_j = \lfloor D/\log q_j \rfloor$, i.e. $D - \log q_j \leq d_j \log q_j \leq D$. But in fact, for $q_1 \leq \cdots \leq q_m$ and q_m fixed, we have $\lim_{D\to\infty} \frac{d_j}{D/\log q_j} = 1$. Hence for D large enough, d_j and $D/\log q_j$ are very close to each other and in later estimates, it suffices to use $D/\log q_j$. We will write* $\boldsymbol{d_j \sim D/\log q_j}$ ***for** D **large enough** for this.*

3.4 Zero estimates: Roth's Lemma

In this section, we state Roth's Lemma, use it to prove Lemma 3.3.2 (Step 2 of the proof of Roth's Theorem), and prove Roth's Lemma.

Lemma 3.4.1 (Roth's Lemma)**.** *Let $P \in \overline{\mathbb{Q}}[x_1,\ldots,x_m]$, not identically zero, of partial degrees at most $d_1,\ldots,d_m$ and $d_j \geq 1$. Let $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_m) \in \overline{\mathbb{Q}}^m$ and let $0 < \sigma \leq \frac{1}{2}$. Assume that*

(i) the weights $d_1,\ldots,d_m$ are rapidly decreasing, i.e.

$$d_{j+1}/d_j \leq \sigma;$$

(ii) the point $(\xi_1,\ldots,\xi_m)$ has components with large height, i.e.

$$\min_j d_j h(\xi_j) \geq \sigma^{-1}(h(P) + 4md_1).$$

Then we have

$$\operatorname{ind}(P;\mathbf{d};\boldsymbol{\xi}) \leq 2m\sigma^{1/2^{m-1}}. \tag{3.4.1}$$

3.4.1 Proof of Lemma 3.3.2 by Roth's Lemma

We will apply Roth's Lemma to the polynomial P constructed in §3.3.1 (Step 1 of the proof of Roth's Theorem) and $\boldsymbol{\xi} = (p_1/q_1,\ldots,p_m/q_m)$. Let us explain the parameters.

Fix $\sigma = \epsilon'^{2^{m-1}} \in (0,1/2]$ (recall our choice $\epsilon' \in (0,1/6)$ in Step 0 of the proof of Roth's Theorem).

Recall our choices $d_j = \lfloor D/\log q_j \rfloor \sim D/\log q_j$ for D large enough and $\log q_{j+1} \geq M \log q_j$. Thus hypothesis (i) of Roth's Lemma is verified if we set

$$M \geq 2\sigma^{-1} \quad \text{and} \quad D \text{ large enough.} \tag{3.4.2}$$

Next, using $d_j h(p_j/q_j) \geq d_j \log q_j \sim D$, $d_m \leq \cdots \leq d_1 \leq D/\log q_1 < D/L$ and the height bound on P given by (3.3.3), we see that hypothesis (ii) of Roth's Lemma is verified if we set

$$D \geq \sigma^{-1}(C+4)m\frac{D}{L}$$

with C the constant depending only on α and m from (3.3.3).

Now we choose M and D as in (3.4.2) and $L \geq \sigma^{-1}(C+4)m$. Then we can apply Roth's Lemma to P and $\boldsymbol{\xi} = (p_1/q_1, \ldots, p_m/q_m)$ to get $\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\xi}) \leq 2m\sigma^{1/2^{m-1}} = 2m\epsilon'$. So there exists $\boldsymbol{\mu}$ such that $\partial_{\boldsymbol{\mu}} P(\boldsymbol{\xi}) \neq 0$ and $\sum_{j=1}^{m} \frac{\mu_j}{d_j} \leq 2m\epsilon'$.

We claim that $Q := \partial_{\boldsymbol{\mu}} P$ is what we desire. Let us check the conclusions for Lemma 3.3.2. Part (ii) is done. For part (i), it suffices to apply Lemma 3.2.3 (iii), the construction $\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) \geq (1/2 - \epsilon')m$ for P and $\sum_{j=1}^{m} \frac{\mu_j}{d_j} \leq 2m\epsilon'$. For (iii), we use Lemma 3.2.1 and the height bound on P (3.3.3) to get

$$h(Q) = h(\partial_{\boldsymbol{\mu}} P) \leq h(P) + (\log 2)\sum d_j \leq C_1 \sum d_j$$

where C depends only on α and m, when all $d_j \to \infty$. Again by using $d_j \log q_j \sim D$ and $\log q_j \geq \log q_1 > L$, we can conclude. □

3.4.2 Proof of Roth's Lemma

We prove Roth's Lemma by induction on m. Notice that for the base step $m = 1$, we in fact prove a stronger bound.

For the base step $m = 1$, we will prove the better bound

$$\operatorname{ind}(P; d_1; \xi_1) \leq \sigma. \tag{3.4.3}$$

By definition of the index, we have that $(x_1 - \xi_1)^{\operatorname{ind}(P;d_1;\xi_1)d_1}$ divides P. Thus we can apply Theorem 1.3.4 to get

$$h(P) \geq -d_1 \log 2 + \operatorname{ind}(P; d_1; \xi_1)d_1 \cdot h(x_1 - \xi_1) \geq -d_1 \log 2 + \operatorname{ind}(P; d_1; \xi_1)d_1 \cdot h(\xi_1).$$

Thus

$$\operatorname{ind}(P; d_1; \xi_1) \leq (h(P) + d_1 \log 2)/d_1 h(\xi_1) \leq \sigma.$$

So we are done for the base step. Notice that hypothesis (ii) for $m = 1$ can be weakened to be $d_1 h(\xi_1) \geq \sigma^{-1}(h(P) + \log 2 \cdot d_1)$.

Now we do the induction step. Assume that Roth's Lemma is proved for $1, \ldots, m-1$. We wish to prove it for m.

We will use the *Wronskian criterion* for linear independence.

Proposition 3.4.2. *Let $\varphi_1, \ldots, \varphi_n$ be polynomials in $\overline{\mathbb{Q}}[x_1, \ldots, x_m]$. Then $\varphi_1, \ldots, \varphi_n$ are linearly independent over $\overline{\mathbb{Q}}$ if and only if some generalized Wronskian*

$$W_{\boldsymbol{\mu_1}, \ldots, \boldsymbol{\mu_n}}(x_1, \ldots, x_m) := \det \begin{pmatrix} \partial_{\boldsymbol{\mu}_1}\varphi_1 & \partial_{\boldsymbol{\mu}_1}\varphi_2 & \cdots & \partial_{\boldsymbol{\mu}_1}\varphi_n \\ \partial_{\boldsymbol{\mu}_2}\varphi_1 & \partial_{\boldsymbol{\mu}_2}\varphi_2 & \cdots & \partial_{\boldsymbol{\mu}_2}\varphi_n \\ \cdot & \cdot & \cdots & \cdot \\ \partial_{\boldsymbol{\mu}_n}\varphi_1 & \partial_{\boldsymbol{\mu}_n}\varphi_2 & \cdots & \partial_{\boldsymbol{\mu}_n}\varphi_n \end{pmatrix},$$

with $|\boldsymbol{\mu}_i| = \mu_1^{(i)} + \mu_2^{(i)} + \cdots + \mu_m^{(i)} \leq i - 1$, is not identically zero.

We will finish the proof of Roth's Lemma assuming Proposition 3.4.2. To perform the splitting of the Wronskian, we write the polynomial $P \in \overline{\mathbb{Q}}[x_1, \dots, x_m]$ in the form

$$P = \sum_{j=0}^{s} f_j(x_1, \dots, x_{m-1}) g_j(x_m)$$

with $s \leq d_m$ and where the f_j's (similarly the g_j's) are linearly independent polynomials over $\overline{\mathbb{Q}}$.

Set

$$U(x_1, \dots, x_{m_1}) := \det(\partial_{\boldsymbol{\mu}_i} f_j)_{i,j=0,\dots,s}$$

with $\boldsymbol{\mu}_i = (\mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_{m-1}^{(i)})$ such that $|\boldsymbol{\mu}_i| \leq s \leq d_m$, and

$$V(x_m) := \det(\partial_\nu g_j)_{\nu,j=0,\dots,s}.$$

By Proposition 3.4.2, we may choose such U and V that they are both not identically 0. Set

$$W(x_1, \dots, x_m) := \det(\partial_{\boldsymbol{\mu}_i,\nu} P) = U(x_1, \dots, x_{m_1}) V(x_m).$$

We wish to apply the induction hypothesis to U and V. Thus we need to analyse the their degrees and heights.

For degrees, it is easy to see that the partial degrees of U are at most $(s+1)d_1, \dots, (s+1)d_{m-1}$, and $\deg V \leq (s+1)d_m$.

Since $d_{j+1}/d_j \leq \sigma \leq 1/2$ by hypothesis (i), we have $d_1 + \ldots + d_m \leq 2d_1$.

For heights, Theorem 1.3.4 yields $h(W) \geq h(U) + h(V) - (s+1)(d_1 + \cdots + d_m)\log 2 \geq h(U) + h(V) - (s+1)(2\log 2)d_1 \geq h(U) + h(V) - (s+1)d_1$. We claim that

$$h(W) \leq (s+1)(h(P) + 3d_1). \tag{3.4.4}$$

Indeed, by expansion, the determinant W is a sum of $(s+1)!$ terms, each of which is the product of $s+1$ polynomials of the form $\partial_{\boldsymbol{\mu}_i,\nu} P$ for some $\boldsymbol{\mu}_i$ and ν. Thus by the proof of Proposition 1.3.12, Theorem 1.3.4 and Lemma 3.2.1, we have[5]

$$h(W) \leq (s+1)\left(h(P) + (d_1 + \ldots + d_m)\log 2\right) + (d_1 + \ldots + d_m)\log 2 + \log(s+1)!.$$

Hence we can establish (3.4.4) because $d_1 + \ldots + d_m \leq 2d_1$ and $\log(s+1)! \leq (s+1)\log(s+1) \leq (s+1)\log(d_m+1) \leq (s+1)d_m \leq (s+1)d_1/2$.

From the previous paragraph, we can conclude $h(U) \leq (s+1)(h(P) + 4d_1)$ and $h(V) \leq (s+1)(h(P) + 4d_1)$, because both heights are non-negative by definition. Now hypothesis (ii) of Roth's Lemma implies

$$\min_j (s+1)d_j h(\xi_j) \geq \sigma^{-1}(h(U) + 4(m-1)(s+1)d_1) \quad \text{and} \quad (s+1)d_m h(\xi_m) \geq \sigma^{-1}(h(V) + 4(s+1)d_m).$$

So we can apply the induction hypothesis to U, $((s+1)d_1, \dots, (s+1)d_{m-1})$ and $(\xi_1, \dots, \xi_{m-1})$ (resp. to V, $(s+1)d_m$ and ξ_m) to get

$$\operatorname{ind}(U; (d_1, \dots, d_{m-1}); (\xi_1, \dots, \xi_{m-1})) \leq 2(m-1)(s+1)\sigma^{1/2^{m-2}} \quad \text{and} \quad \operatorname{ind}(V; d_m; \xi_m) \leq (s+1)\sigma. \tag{3.4.5}$$

[5] One cannot directly apply Proposition 1.3.12 here. Instead, one goes into its proof, which is essentially the proof of Proposition 1.2.8. Notice that all the $\|x_j^{(k)}\|_v$'s at the end of that proof has the same upper bound in terms of P (because they are all derivatives of P), so in the long inequalities at the of that proof there is not need to take the sum $\sum_{1 \leq k \leq r}$.

Here for V, we have used the better bound obtained in the base step $m = 1$. Therefore

$$\mathrm{ind}(W; \mathbf{d}; \boldsymbol{\xi}) = \mathrm{ind}(U; (d_1, \ldots, d_{m-1}); (\xi_1, \ldots, \xi_{m-1})) + \mathrm{ind}(V; d_m; \xi_m) \leq 2(m-1)(s+1)\sigma^{1/2^{m-2}} + (s+1)\sigma. \quad (3.4.6)$$

It remains to relate the index of P with the index of W. To ease notation, we use $\mathrm{ind}(\cdot)$ to denote $\mathrm{ind}(\cdot; \mathbf{d}; \boldsymbol{\xi})$. For each $\boldsymbol{\mu}_i$ and ν, Lemma 3.2.3(iii) yields

$$\begin{aligned}
\mathrm{ind}(\partial_{\boldsymbol{\mu}_i,\nu} P) &\geq \mathrm{ind}(P) - \sum_{j=1}^{m-1} \frac{\mu_j^{(i)}}{d_j} - \frac{\nu}{d_m} \\
&\geq \mathrm{ind}(P) - \frac{d_m}{d_{m-1}} - \frac{\nu}{d_m} \quad \text{since } \mu_1^{(i)} + \ldots + \mu_{m-1}^{(i)} \leq i - 1 \leq s \leq d_m \\
&\geq \mathrm{ind}(P) - \frac{\nu}{d_m} - \sigma.
\end{aligned}$$

This bound can be automatically improved since the index is always non-negative. So

$$\mathrm{ind}(\partial_{\boldsymbol{\mu}_i,\nu} P) \geq \max\left\{\mathrm{ind}(P) - \frac{\nu}{d_m}, 0\right\} - \sigma.$$

Again, we expand the determinant W. We can write W explicitly in the following way: $W = \sum_\pi \prod_{i=0}^s \partial_{\boldsymbol{\mu}_i,\pi(i)} P$ with π running over all permutation of the set $\{0, \ldots, s\}$. Thus we can apply parts (i) and (ii) of Lemma 3.2.3 to get $\mathrm{ind}(W) \geq \min_\pi \left(\sum_{i=0}^s \mathrm{ind}(\partial_{\boldsymbol{\mu}_i,\pi(i)} P)\right)$. So we have

$$\begin{aligned}
\mathrm{ind}(W) &\geq \min_\pi \sum_{i=0}^s \left(\max\left\{\mathrm{ind}(P) - \frac{\pi(i)}{d_m}, 0\right\} - \sigma\right) \\
&= \sum_{i=0}^s \left(\max\left\{\mathrm{ind}(P) - \frac{i}{d_m}, 0\right\} - \sigma\right) \\
&\geq (s+1)\min\left\{\frac{1}{2}\mathrm{ind}(P), \frac{1}{2}\mathrm{ind}(P)^2\right\} - (s+1)\sigma
\end{aligned}$$

where the last step comes from $s \leq d_m$ and the elementary inequality

$$\sum_{i=0}^s \max\left\{t - \frac{i}{s}, 0\right\} \geq (s+1)\min\left\{\frac{1}{2}t, \frac{1}{2}t^2\right\}.$$

Combined with (3.4.6), this lower bound of $\mathrm{ind}(W)$ yields

$$\min\{\mathrm{ind}(P), \mathrm{ind}(P)^2\} \leq 4(m-1)\sigma^{1/2^{m-2}} + 2\sigma.$$

But $\mathrm{ind}(P) \leq m$ by definition. So we have

$$\mathrm{ind}(P)^2 \leq m\left(4(m-1)\sigma^{1/2^{m-2}} + 2\sigma\right) \leq 4m^2\sigma^{1/2^{m-2}}.$$

Hence we are done. □

3.4.3 Proof of Proposition 3.4.2

We start with $\Leftarrow$. Assume $\varphi_1, \ldots, \varphi_n$ are linearly dependent over $\overline{\mathbb{Q}}$. Then all generalized Wronskians vanish. Indeed, we have $c_1\varphi_1 + \cdots + c_n\varphi_n = 0$ for some $c_1, \ldots, c_n \in \overline{\mathbb{Q}}$ not all zero. Applying the operators $\partial_{\boldsymbol{\mu}_i}$ to this relation, we obtain a linear system in the coefficients

c_j and its determinant mush vanish. This determinant is precisely the generalized Wronskian $W_{\boldsymbol{\mu_1},\ldots,\boldsymbol{\mu_n}}(x_1,\ldots,x_m)$.

Let us prove $\Rightarrow$. Assume $\varphi_1,\ldots,\varphi_n$ are linearly independent over $\overline{\mathbb{Q}}$

We assume the following lemma, which is a particular case of the proposition but itself is a classical result.

Lemma 3.4.3. *Let $f_1,\ldots,f_n \in \overline{\mathbb{Q}}[t]$ be n polynomials in 1 variable. Then $f_1,\ldots,f_n$ are linearly independent over $\overline{\mathbb{Q}}$ if and only if the Wronskian*

$$W(t) := \det\left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{i-1} f_j\right)_{1\le i,j\le n}$$

is not identically zero.

We will reduce Proposition 3.4.2 to the situation of this lemma by using the *Kronecker substitution* which we have seen in the proof of Gauß's Lemma.

Fix an integer d which is large than the partial degrees of the φ_j's. Set $x_j := t^{d^{j-1}}$ for $j \in \{1,\ldots,n\}$. Then $\varphi_1,\ldots,\varphi_n$ are linearly independent over $\overline{\mathbb{Q}}$ if and only if the polynomials

$$\Phi_j(t) := \varphi_j(t, t^d, \ldots, t^{d^{m-1}})$$

are linearly independent over $\overline{\mathbb{Q}}$. Thus the lemma above implies that the polynomial

$$W(t) = \det\left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{i-1} \Phi_j\right)_{1\le i,j\le n}$$

is not identically 0. But

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{i-1} \Phi_j = \sum_{|\boldsymbol{\mu}|\le i-1} a_{\boldsymbol{\mu},i}(t;d,m)\partial_{\boldsymbol{\mu}}\varphi_j(t,t^d,\ldots,t^{d^{m-1}})$$

for some universal polynomials $a_{\boldsymbol{\mu},i}(t;d,m) \in \mathbb{Q}[t]$. Thus $W(t)$ is a linear combination of generalized Wronskians $W_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n}(t,t^d,\ldots,t^{d^{m-1}})$ with $|\boldsymbol{\mu}_i| \le i-1$. Since $W(t)$ is not identically 0, some generalized Wronskian is not identically zero. Hence we are done. □

Proof of Lemma 3.4.3. The direction $\Leftarrow$ is easy. Let us prove the direction $\Rightarrow$ by induction on n. The base step $n = 1$ is clearly true.

Assume $\Rightarrow$ is proved for $1,\ldots,n-1$. For n and the polynomials $f_1,\ldots,f_n$, assume that $W(t)$ is identically 0. For each $j \in \{1,\ldots,n\}$, set $W_j(t)$ to be the Wronskian of the $n-1$ polynomials by omitting f_j. Then by expanding the determinant $W(t)$ by the last row, we get $W(t) = \sum_{j=1}^n W_j \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n-1} f_j = \sum_{j=1}^n W_j f_j^{(n-1)}$. Here we change the notation and denote by $f_j^{(i)}$ the i-th derivative of f_j. Thus

$$W_1 f_1^{(n-1)} + \cdots + W_n f_n^{(n-1)} \equiv 0.$$

We claim that $W_1 f_1 + \cdots + W_n f_n \equiv 0$. Indeed, the left hand side is the determinant of the $n \times n$-matrix $\begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ \cdot & \cdot & \cdots & \cdot \\ f_1^{(n-2)} & f_2^{(n-2)} & \cdots & f_n^{(n-2)} \\ f_1 & f_2 & \cdots & f_n \end{pmatrix}$, by the expansion along the last row. Similarly we have $\sum_j W_j f_j^{(i)} \equiv 0$

for each $i \in \{1, \dots, n-2\}$. Thus we obtain a system of n equalities of polynomials

$$\begin{aligned} W_1 f_1 + \cdots + W_n f_n &\equiv 0 \\ W_1 f_1' + \cdots + W_n f_n' &\equiv 0 \\ &\cdots \\ W_1 f_1^{(n-1)} + \cdots + W_n f_n^{(n-1)} &\equiv 0 \end{aligned}$$

Differentiating each of the first $n-1$ equality and subtracting the next following one, we get the following new system

$$\begin{aligned} W_1' f_1 + \cdots + W_n' f_n &\equiv 0 \\ W_1' f_1' + \cdots + W_n' f_n' &\equiv 0 \\ &\cdots \\ W_1' f_1^{(n-1)} + \cdots + W_n' f_n^{(n-1)} &\equiv 0 \end{aligned}$$

Next multiplying the i-th equality ($i = 1, 2, \dots, n-1$) by the minor of W_n corresponding to $f_1^{(i-1)}$ and adding the equalities thus obtained together, we get

$$W_1' W_n - W_1 W_n' \equiv 0.$$

If $W_1 \equiv 0$, then $f_2, \dots, f_n$ are linearly dependent over $\overline{\mathbb{Q}}$ by induction hypothesis, and so are $f_1, \dots, f_n$. Suppose $W_1 \not\equiv 0$. Then we can divide both sides by W_1^2 (notice that W_1 is a polynomial and hence has only finitely many zeros) and get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{W_n}{W_1}\right) \equiv 0.$$

Thus $W_n \equiv c_1 W_1$ for some constant $c_1 \in \overline{\mathbb{Q}}$. Similarly we have $W_n \equiv c_j W_j$ for each $j \in \{2, \dots, n-1\}$ or the conclusion already holds true. Thus either the conclusion holds true, or

$$W_n(c_1 f_1 + \ldots + c_{n-1} f_{n-1} + f_n) \equiv 0.$$

Again either $W_n \equiv 0$ (and hence the conclusion holds true), or $c_1 f_1 + \ldots + c_{n-1} f_{n-1} + f_n \equiv 0$ (and hence the conclusion holds true)[6]. So in either case we are done for the induction step. □

3.5 An alternative approach to the zero estimates: Dyson's Lemma

In this section, we explain an alternative approach to the zero estimates.

In the proof of Roth's Theorem presented in previous sections of this chapter, we used Roth's Lemma (Lemma 3.4.1) to do the zero estimates and found a polynomial P having large index at $\boldsymbol{\alpha} = (\alpha, \dots, \alpha)$ but small index at $(p_1/q_1, \dots, p_m/q_m)$. Roth's Lemma is *arithmetic* in nature: the polynomial P has coefficients in $\overline{\mathbb{Q}}$, we are interested in its order of vanishing at an algebraic point, and a hypothesis (hypothesis (ii)) on the given data is about the heights.

An alternative approach to establish the small index of $P(p_1/q_1, \dots, p_m/q_m)$, developed by Esnault–Viehweg building upon previous work of Dyson, Bombieri and Viola, is the so-called *Dyson's Lemma*. It is a geometric approach (and hence works over any algebraically closed field of characteristic 0) and the philosophy is as follows. Suppose that whichever P we have constructed with large index at $\boldsymbol{\alpha}$ also has large index at $(p_1/q_1, \dots, p_m/q_m)$. Then certain linear conditions on the space of all polynomials of partial degree $d_1, \dots, d_m$ fail to be independent. Thus in order to get a contradiction, it suffices to establish this independence.

[6] Notice that the zeros of W_n are isolated if $W_n \not\equiv 0$.

To state Dyson's Lemma, recall the notation $\mathcal{V}_m(t) := \{\mathbf{x} \in \mathbb{R}^m : x_1 + \cdots + x_m \leq t,\ 0 \leq x_j \leq 1\}$ and $V_m(t)$ the volume of $\mathcal{V}_m(t)$ with respect to the usual Lebesgue measure on $\mathbb{R}^m$. We set $V_m(t) = 0$ for $t < 0$. The arithmetic meaning of $V_m(t)$ was explained in the proof of Lemma 3.2.4: *In the linear system related to constructing a polynomial of index $\geq t$ at a given point (with respect to the partial degrees $d_1, \ldots, d_m$), $d_1 \cdots d_m V_m(t)$ is asymptotically the number of equations.*

Theorem 3.5.1 (Dyson's Lemma)**.** *Let $\mathbf{d} = (d_1, \ldots, d_m)$ be such that $d_1 \geq d_2 \geq \cdots d_m \geq 1$ are positive integers.*

Let $\boldsymbol{\zeta}_1 = (\zeta_1^{(1)}, \ldots, \zeta_m^{(1)}), \ldots, \boldsymbol{\zeta}_{r+1} = (\zeta_1^{(r+1)}, \ldots, \zeta_m^{(r+1)})$ be $r+1$ points in $\mathbb{C}^m$ such that $\zeta_k^{(i)} \neq \zeta_k^{(j)}$ for all $k \in \{1, \ldots, m\}$ and all $i \neq j$.[7]

Let $P \in \mathbb{C}[x_1, \ldots, x_m]$ of partial degrees at most $d_1, \ldots, d_m$, and denote by $t_i := \mathrm{ind}(P; \mathbf{d}; \boldsymbol{\zeta}_i)$ for all $i \in \{1, \ldots, r+1\}$. Then we have

$$\sum_{i=1}^{r+1} V_m(t_i) \leq \prod_{j=1}^{m} \left(1 + (r'-2) \sum_{l=j+1}^{m} \frac{d_l}{d_j}\right) \tag{3.5.1}$$

where $r' := \max\{r+1, 2\}$.

The field $\mathbb{C}$ in the statement can be replaced by any algebraically closed field of characteristic 0.

We will not prove Theorem 3.5.1, but only see how Theorem 3.5.1 can be used to prove Roth's Theorem.

We need the following technical lemma.

Lemma 3.5.2. *Let $r \geq 2$ be an integer and let $\epsilon' > 0$. Then there exists an integer $m_0 = m_0(r, \epsilon') \geq 2$ with the following property. For all $m \geq m_0$, there exist a real number $\tau > 1$ such that*

$$rV_m(\tau) < 1 < rV_m(\tau) + V_m(1) \quad \text{and} \quad (2+\epsilon')(\tau - 1) > m. \tag{3.5.2}$$

Proof. We prove the lemma by taking τ such that

$$rV_m(\tau) = 1 - \frac{1}{2m!}.$$

Indeed, such a τ exists, and the first inequality in (3.5.2) holds true because $V_m(1) = 1/m!$.

Let us prove $(2+\epsilon')(\tau-1) > m$. We start by trying to solve the inequality

$$\sqrt{\frac{\log r - \log\left(1 - \frac{1}{2m!}\right)}{6m}} + \frac{1}{m} < \frac{1}{2} - \frac{1}{2+\epsilon'}.$$

Since the left hand side tends to 0 as $m \to \infty$, there exists an integer $m_0 \geq 2$ such that this inequality holds true for all $m \geq m_0$. Let us show that $(2+\epsilon')(\tau-1) > m$ for all these m. Recall $V_m((1/2-\eta)m) \leq e^{-6m\eta^2}$ by Lemma 3.2.5, for all $0 \leq \eta \leq 1/2$. Take η such that $(1/2-\eta)m = \tau$. Then we have

$$\eta \leq \sqrt{\frac{\log r - \log\left(1 - \frac{1}{2m!}\right)}{6m}} < \frac{1}{2} - \frac{1}{2+\epsilon'} - \frac{1}{m}.$$

So

$$\frac{\tau - 1}{m} = \frac{1}{2} - \eta - \frac{1}{m} > \frac{1}{2+\epsilon'}.$$

This yields $(2+\epsilon')(\tau-1) > m$. We are done. □

[7] Namely, if we look at the projection to the k-th component, then we still get $r+1$ different points in $\mathbb{C}$.

Now let us sketch the proof of Roth's Theorem by using Dyson's Lemma instead of Roth's Lemma.

Proof of Theorem 3.3.1. Let $\alpha \in \mathbb{R}$ and $\epsilon > 0$ be as in Roth's Theorem. Assume that there are infinitely rational approximations. Then for each m, L and M, we can find rational approximations p_j/q_j ($j \in \{1, \dots, m\}$ and $q_j \geq 1$), *i.e.* $|\alpha - p_j/q_j| \leq q_j^{-(2+\epsilon)}$, such that they are (L, M)-independent, *i.e.* $\log q_1 > L$ and $\log q_{j+1} > M \log q_j$ for each j. This is the same as Step 0.

Now let us do Step 1, *i.e.* construct an auxiliary polynomial P of large index at $\boldsymbol{\alpha}$ and of small height.

Set $r = [\mathbb{Q}(\alpha) : \mathbb{Q}]$. Write $\alpha_1 = \alpha$, $\alpha_2, \dots, \alpha_r$ for the Galois conjugates of α.

Let $\epsilon' > 0$, m and τ be from Lemma 3.5.2. Then

$$\tau - 1 > \frac{m}{2 + \epsilon'}.$$

Take another parameter D, and set $d_j = \lfloor D/\log q_j \rfloor$ for each j.

By Lemma 3.2.4 and the choice that $rV_m(\tau) < 1$, there exists a polynomial $P \in \mathbb{Z}[x_1, \dots, x_m]$ of large index at $\boldsymbol{\alpha}$ and of small height. More precisely,

(i) $\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) \geq \tau$;

(ii) As $d_j \to \infty$ for all $j \in \{1, \dots, m\}$, we have

$$h(P) \leq C \cdot 2m!(d_1 + \cdots + d_m) < C \cdot 2m! \frac{mD}{L} \tag{3.5.3}$$

with C a suitable constant depending only on α and m.

Condition (i) is equivalent to: For each $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$ with $\sum \frac{\mu_j}{d_j} < \tau$, we have $\partial_{\boldsymbol{\mu}} P(\boldsymbol{\alpha}) = 0$. Since P has integer coefficients, applying the Galois action yields $\partial_{\boldsymbol{\mu}} P(\boldsymbol{\alpha}_j) = 0$ for each $j \in \{1, \dots, r\}$ and each such $\boldsymbol{\mu}$, where $\boldsymbol{\alpha}_j = (\alpha_j, \dots, \alpha_j)$. Hence $\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}_j) \geq \tau$ for all $j \in \{1, \dots, r\}$.

Now we use Dyson's Lemma to accomplish Step 2 (non-vanishing at the rational point).

Choose the parameter M in the following way: by Lemma 3.5.2, we can find an $M \gg 1$ such that

$$rV_m(\tau) + V_m(1) > \prod_{j=1}^{m} \left(1 + (r-1) \sum_{l=j+1}^{m} \frac{1}{M^{l-j}}\right). \tag{3.5.4}$$

Since $\log q_{j+1} > M \log q_j$ for all j and $d_j \sim D/\log q_j$ for D large enough, the inequality above can be translated into (for sufficiently large D)

$$rV_m(\tau) + V_m(1) > \prod_{j=1}^{m} \left(1 + (r-1) \sum_{l=j+1}^{m} \frac{d_l}{d_j}\right). \tag{3.5.5}$$

Apply Dyson's Lemma (Theorem 3.5.1) to the points $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r, \boldsymbol{\xi} := (p_1/q_1, \dots, p_m/q_m)$. Then we get

$$rV_m(\tau) + V_m(\operatorname{ind}(P; \mathbf{d}; \boldsymbol{\xi})) \leq \prod_{j=1}^{m} \left(1 + (r-1) \sum_{l=j+1}^{m} \frac{d_l}{d_j}\right). \tag{3.5.6}$$

Comparing (3.5.5) and (3.5.6), we get

$$\mathrm{ind}(P;\mathbf{d};\boldsymbol{\xi}) < 1.$$

Take $\boldsymbol{\mu}$ be such that $\partial_{\boldsymbol{\mu}} P(\boldsymbol{\xi}) \neq 0$ and that $\sum \frac{\mu_j}{d_j} = \mathrm{ind}(P;\mathbf{d};\boldsymbol{\xi}) < 1$. Set $Q = \partial_{\boldsymbol{\mu}} P$. Then

(i) $\mathrm{ind}(Q;\mathbf{d};\boldsymbol{\alpha}) \geq \mathrm{ind}(P;\mathbf{d};\boldsymbol{\alpha}) - \sum \frac{\mu_j}{d_j} > \tau - 1 > \frac{m}{2+\epsilon'}$;

(ii) $Q(p_1/q_1, \ldots, p_m/q_m) \neq 0$;

(iii) $h(Q) \leq C' \cdot 2m! \frac{mD}{L}$.

Here (i) uses Lemma 3.2.3,(iii), and (iii) uses Lemma 3.2.1.

Then one repeats the argument as in Step 3, 4 and 5 of §3.3 and eventually get

$$1 \geq \frac{2+\epsilon}{2+\epsilon'} - \frac{C' \cdot 2m!}{L} - \frac{(m+1)\log 2}{mD}.$$

This gives a contradiction by letting $\epsilon' \to 0$, $L \to \infty$ and $D \to \infty$. □