Universitat de Barcelona

Facultat de Matemàtiques
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

Diophantine Approximation in the framework of Roth's Theorem

Autor: Eduard Valcarce Dalmau

Director: Dr. Martín Sombra
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, January 24, 2023

Abstract

The main goal of this work is to understand a proof of a generalized version of Roth's theorem proposed by Lang. Due to the large scope of this proof, we will begin with older, more foundational results in Diophantine approximation, as they provide context, and introduce the general structure of the main proof in this work.

Then we will study the theory of absolute values over number fields, in order to use the results and tools derived from it, such as the height functions. These functions, together with the index of a polynomial will play a huge role in the proof of the more general version of Roth's theorem.

We will then present the proof of the theorem, and finish off this work with a few applications of the theorem, as well as a discussion on an inherent limitation of the proof that carries over into other renowned theorems that depend on Roth's theorem, such as Falting's theorem on the finiteness of rational points in curves of genus greater or equal to two.

2020 Mathematics Subject Classification. 11J68, 11D45, 11D57, 11J17, 11J25, 11J82,

Acknowledgements

First of all, I would like to thank my tutor Dr. Martín Sombra for helping me narrow down the subject of this project, and supplying me with many books afterwards, he was always there when I had any query, and helped me look for alternative proofs for results if it was necessary. Thanks to him I have learned about this topic which had piqued my curiosity for years.

I would also like to thank my parents for supporting me and being there when I needed them or simply wanted to have a chat. Just as important was my cat Luke, who always makes me smile and forces me to take a break every now and then.

Lastly, I want to thank my friends from university, especially the ones from the double degree: they made these long years pass by incredibly quickly, and made going to classes a joy.
Thank you all.

Agraïments

Primer de tot, voldria agrair al meu tutor, el Dr. Martín Sombra per ajudar-me a concretar el tema del TFG i deixar-me llibres per llegir sobre el tema, sempre m'ha estat ajudant amb dubtes i m'ajudava a buscar demostracions alternatives si era necessari. Gràcies a ell he après molt sobre aquest tema que em portava anys intrigant.

També m'agradaria agrair als meus pares per donar-me suport i per estar allà tot i que nomes fos per xerrar. Igual d'important és el meu gat Luke, que sempre em posa un somriure a la cara i em fa prendre descansos ocasionalment.

Finalment, m'agradaría agrair als meus amics de la universitat, sobre tot els del doble grau: han fet que aquests anys passin molt ràpid, i feien que anar a classe fos entretingut.
Gràcies a tots.

Contents

1 Introduction 1

2 Classical results in Diophantine approximation 3

3 Theory of absolute values and absolute valued number fields 6

3.1 Introduction and the degree formula 6

3.2 The product formula 8

3.3 Heights 11

4 Preliminary results 14

4.1 Notation and defintions 14

4.2 Results 14

5 Proof of Roth's Theorem 24

5.1 Construction of the Auxiliary Polynomial 24

5.2 The index of the Auxiliary Polynomial must be large 25

5.3 The index of the Auxiliary Polynomial must be small 29

5.4 Completion of the Proof of Roth's Theorem 39

5.5 On the effectivity of the proof 42

6 Applications of Roth's theorem: the unit equation 43

1 Introduction

Fundamentally, Diophantine approximation is about how well an irrational number can be approximated by rational numbers. It is a well known fact that $\mathbb{Q}$ is dense in $\mathbb{R}$, so obviously any irrational number can be approximated by rational numbers to any given accuracy, which is usually written as $\forall \alpha \in \mathbb{R}\ \forall \varepsilon > 0$, there exist $p/q \in \mathbb{Q}$ such that $|\alpha - p/q| < \varepsilon$. Our intuition tells us that to make $|\alpha - p/q|$ small, we should make p and q big, to, crudely, approximate α to more decimal places. We encode this intuition into the statement

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{f(q)}$$

for some increasing function f. We do not include p on the right side of the inequality because it is in fact q in this case that limits how precise p/q can be. Moreover, p is closely related to q through the fact that they approximate α.

An obvious question is how the number of approximations of α is affected by the function f. Clearly if f is a positive constant, there are an infinite number of approximations, but it is not at all clear whether this remains the case for a function that increases 'more rapidly' with q. In fact, if one were to take this reasoning to it's natural limit and take $1/f(q) = 0$, then there are obviously no integers p, q that satisfy the inequality for $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. This suggests the possibility that just by knowing α and f, there may be a way to know whether or not there are finitely many solutions to the inequality, perhaps even the exact number of them.

In 1842 Dirichlet was one of the first notable mathematicians to provide insight into this problem, by finding a lower bound for $f(q)$ that preserves an infinite number of solutions to the inequality. Dirichlet proved that given $\alpha \in \mathbb{R}, \forall \varepsilon > 0$ there are infinitely many $p/q \in \mathbb{Q}$ such that

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}$$

and shortly thereafter, in 1844, Liouville would provide an upper bound for $f(q)$, by proving that given an algebraic number $\alpha \in \mathbb{R}$ of degree d over $\mathbb{Q}$, there exist only finitely many $p/q \in \mathbb{Q}$ such that

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^d}.$$

These results would set the tone for the future study of this inequality, which is why the results that follow are improvements on the exponent d in Liouville's proof. Mathematicians were looking for an exponent that determined whether the inequality had finitely or infinitely many solutions. For a long time this conjectured exponent seemed to be a function of α, $\tau(\alpha)$, as the upper bounds kept getting stronger but always kept a dependence with d.

Liouville	1844	$\tau(\alpha) \leq d$
Thue	1909	$\tau(\alpha) \leq 1 + d/2$
Siegel	1921	$\tau(\alpha) \leq 2\sqrt{d}$
Gelfand, Dyson	1947	$\tau(\alpha) \leq \sqrt{2d}$

In 1955, however, Roth proved his major result, Roth's Theorem [9], which proved that given an algebraic number $\alpha \in \overline{\mathbb{Q}}$. $\forall \varepsilon > 0$, there are only finitely many $p/q \in \mathbb{Q}$ such that

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

Together with Liouville's result, this completely characterized the relationship between the aforementioned $\tau(\alpha)$ and the solutions to $|\alpha - p/q| < 1/q^{\tau(\alpha)}$.

The importance of this result cannot be understated, as evidenced by the numerous generalizations it has gotten over the years. In fact, the centerpiece of this work is the proof of a slight modification of Lang's generalization of Roth's Theorem [8]:

Theorem 1.1 (Roth's theorem)**.** Let K be a number field, let $S \subset M_K$ be a finite set of absolute values on K, and assume that each absolute value in S has been extended in some way to $\overline{K}$. Given $\alpha \in \overline{K}$, then, $\forall \varepsilon > 0$ there exist only finitely many $\beta \in K$ satisfying the inequality

$$\prod_{v \in S} \min\{1, \|\beta - \alpha\|_v\} \leq \frac{1}{H_K(\beta)^{2+\varepsilon}}. \tag{1}$$

Everything in the previous expression is defined in chapter 3, where we will also explain how this result can be reduced to the one Roth proved in 1955.

In this work we will mainly follow the proof of Roth's Theorem found in [1]. The preliminary results from chapter 3 come mainly from [1, 2, 5]

2 Classical results in Diophantine approximation

Before moving on to the theory of absolute values over number fields, we give the proofs to some of the earliest results in Diophantine approximation. Liouville's result is particularly interesting, as it follows the same general structure as the proof of Roth's Theorem. We will be sure to highlight the similarities.

Proposition 2.1 (Dirichlet, 1842)**.** Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Then there are infinitely many rational numbers $p/q \in \mathbb{Q}$ such that

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}$$

Proof. For any integer Q ≥ 1 consider the set

$$\{q\alpha - \lfloor q\alpha \rfloor |\ q \in \{0, 1, ..., Q\}\}$$

which contains $Q+1$ distinct numbers in $[0,1]$ because α is irrational. Dividing now the interval $[0,1]$ into Q segments of equal length, by the pidgeon hole principle there is at least one of these segments that contains two numbers from the set. This means there exist two integers $0 \leq q_1 < q_2 \leq Q$ such that

$$|(q_1\alpha - \lfloor q_1\alpha \rfloor) - (q_2\alpha - \lfloor q_2\alpha \rfloor)| \leq 1/Q$$

Which given $1 \leq q_2 - q_1 \leq Q$, can be rewritten as

$$\left|\frac{\lfloor q_2\alpha \rfloor - \lfloor q_1\alpha \rfloor}{q_2 - q_1} - \alpha\right| \leq \frac{1}{Q(q_2 - q_1)} \leq \frac{1}{(q_2 - q_1)^2}$$

Relabeling $\lfloor q_2\alpha \rfloor - \lfloor q_1\alpha \rfloor = p \in \mathbb{Z}$ and $q_2 - q_1 = q \in \mathbb{Z}$ we obtain the desired inequality

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}$$

□

Remark 2.1.1 There is a latter result, by Hurwitz in 1891, which proves that the $1/q^2$ from Dirichlet's result can be changed to $1/(q^2\sqrt{5})$, and that $1/\sqrt{5}$ is best possible. More details can be found in [4, Theorems 193 and 194].

Now the first result in the opposite direction, before Roth proved $\tau(\alpha) = 2$.

Proposition 2.2 (Liouville, 1844)**.** Let $\alpha \in \overline{\mathbb{Q}}$ be an algebraic number of degree $d \geq 2$. $\forall \varepsilon > 0$, there are only finitely many $p/q \in \mathbb{Q}$ such that

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^{d+\varepsilon}} \qquad (*)$$

Proof.

Despite being a much weaker result than the one Roth would prove 111 years later, the structure of this proof is remarkably similar. We will show the steps Liouville took to prove this, and we will later see that even in the proof of the more general version of Roth's Theorem, the same steps are taken.

Step 1: Construction of the polynomial
Given we have chosen an algebraic number α with degree d, an obvious candidate is the minimal polynomial of α over $\mathbb{Q}$, which we will call $P(x) \in \mathbb{Z}[X]$.
Step 2: The polynomial must vanish at p/q
We assume that p/q closely approximates α, and want to show that $P(p/q) = 0$. Now, given $P(x) \in \mathbb{Z}[X]$, there is some $N \in \mathbb{Z}$ such that $P(p/q) = \mathrm{N}/q^d$. We now apply Taylor's theorem to expand P around $x = \alpha$, to find an upper bound for $P(p/q)$, which will show $P(p/q)$ must be 0.

$$P(x) = \sum_{i=1}^{d} \frac{1}{i!} \cdot \frac{d^i P}{dx^i}(\alpha)(x-\alpha)^i$$

Notice that the sum ranges from $i = 1$ to $i = d$ because $P(\alpha) = 0$, so for a p/q satisfying (*), this allows us to write

$$\left|\frac{N}{q^d}\right| = |P(p/q)| \leq \left|\alpha - \frac{p}{q}\right| \cdot \left(\sum_{i=1}^{d} \left|\frac{1}{i!} \cdot \frac{d^i P}{dx^i}(\alpha)\right| \cdot |x-\alpha|^{i-1}\right) = C(\alpha)\left|\alpha - \frac{p}{q}\right| \leq \frac{C(\alpha)}{q^{d+\varepsilon}}$$

From the hypothesis that p/q satisfies (*), and where we've labelled the constant term depending only on $\alpha, C(\alpha)$. Rearranging we obtain

$$|N| \leq \frac{C(\alpha)}{q^{\varepsilon}}$$

So taking $q > C(\alpha)^{1/\varepsilon}$ implies that $N \in \mathbb{Z}$ must be 0, which is equivalent to $P(p/q) = 0$.

Step 3: The polynomial cannot vanish at p/q
$P(x)$ is the minimal polynomial of α over $\mathbb{Q}$, meaning it is irreducible in $\mathbb{Q}[X]$, so it follows p/q cannot be a root of $P(x)$.

Despite the simplicity of this step, the proof of Roth's theorem requires the construction of multivariable polynomials, which makes the nonvanishing step the hardest of the whole proof.
Step 4: Proof ot the proposition
We assume that (*) allows infinitely many solutions,to the inequality. This assumption implies that we can choose one of the solutions n/m such that $m > C(\alpha)$, where $C(\alpha)$ is the constant from step 2. We see in step 2 that this implies $P(n/m) = 0$, yet step 3 proves that $P(n/m) \neq 0$. The contradiction implies we can't choose a solution n/m such that $m > c(\alpha)$, which in turn implies we have finitely many solutions n/m to the inequality (*).

□

Remark 2.2.1
This proposition tells us that any given algebraic number has a certain rational number p/q closest to it in the sense of satisfying (*). Liouville used this to explicitly construct a certain class of Transcendental numbers which are now known as *Liouville numbers* due to the following property:
Given $x \in \mathbb{R}$, we say x is a Liouville number if for every $n \in \mathbb{N}$ there exists a rational number $p/q \in \mathbb{Q}$ such that

$$0 < \left|x - \frac{p}{q}\right| < \frac{1}{q^n}.$$

Assuming we have $x \in \mathbb{R}$ that satisfies this inequality yields a contradiction if we assume x is algebraic of degree d over $\mathbb{Q}$, as Liouville's result says there are finitely many $p/q \in \mathbb{Q}$ such that $|x - p/q| < 1/q^d$. Liouville used this to prove for the first time that the number

$$\sum_{i=0}^{\infty} 10^{-n!} = 0.11000100000...$$

now called Liouville's constant is transcendental. This was the first explicit number proven to be transcendental, as Liouville's result was the first one to give a way of constructing transcendental numbers. Even today, aside from numbers like Liouville's constant, constructed specifically to be transcendental, there are very few known transcendental numbers.

Despite these advances brought upon by Liouville's result, it is not good enough for most applications to Diophantine equations, as the exponent $d+\varepsilon$ is often too large. This led to others improving upon the exponent from Liouville's result, as seen in the earlier table, but it was only in 1955 that Roth proved the result for an exponent that didn't depend on the degree, d, of the algebraic number α.

3 Theory of absolute values and absolute valued number fields

Definition 3.1. An absolute value on a field K is a function $|\cdot| : K \to [0, \infty)$ such that

- $|x| \geq 0$, and $|x| = 0$ if and only if $x = 0$
- $|xy| = |x| \cdot |y|$
- $|x + y| \leq |x| + |y|$ (triangle inequality).

Moreover, if an absolute value satisfies the following stronger condition for all $x, y \in K$, too, it is called non-archimedean.

- $|x + y| \leq \max\{|x|, |y|\}$ (Ultrametric inequality)

Now, the distance $d(x, y) := |x-y|$ between any two elements $x, y \in K$ induces a topology on K, and if two absolute values define the same topology, they are called equivalent.

Before the following section, we note the existence of the trivial absolute value:

$$|x| = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases}$$

3.1 Introduction and the degree formula

A *place* v is an equivalence class of non-trivial absolute values acting on a number field K, where two non-trivial absolute values belong to the same equivalence class if and only if they define the same topology over K. The absolute value in the equivalence class determined by the place v is denoted by $|\cdot|_v$. Given the field extension L/K and v a place of K, then any place w of L such that the restriction of $|\cdot|_w$ over K is a representative of v, is said to lie over v, or equivalently, we say that w extends v. This is written as $w|v$, due to the fact that non-archimedean places in number fields correspond to prime ideals.

Given a number field with an absolute value $(K, |\cdot|_K)$, a completion of K, which we denote here by $(L, |\cdot|_L)$, is a number field with an absolute value, complete as a metric space and for which there exists an embedding $i : K \to L$ such that $i(K)$ is dense in L, with $|x|_K = |i(x)|_L \ \forall x \in K$. This completion is unique up to isomorphism.
Now, with our notation, the *completion* of K with respect to a place v is the extension field K_v with a place w such that:

- $w|v$.
- The topology induced by w on K_v is complete.
- K is a dense subset of K_v in that same topology induced by w (In this case the embedding would just be the identity).

The completion exists and is unique up to isometric isomorphisms [3]. Frequently, due to abuse of notation, the place w is denoted v as well.
We have now covered the groundwork for the proposition that will immediately give us the local degree sum formula. Let K be a field with a fixed non.trivial absolute value $|\cdot|_v$.

Proposition 3.2. Let L/K be a finite degree field extension generated by a single element ζ with monic minimal polynomial $f(t)$ over K. Suppose that this polynomial can be decomposed into

$$f(t) = f_1^{n_1}(t) \cdots f_r^{n_r}(t)$$

where $f_i(t) \in K_v[t]$ are different irreducible monic factors, then for each $i \in \{1, 2, ..., r\}$ there exists an injective homomorphism of field extensions over K

$$\begin{aligned} \varphi_i : L &\to K_i \cong K_v[t]/(f_i(t)) \subset \mathbb{C}_v \\ \zeta &\to t \end{aligned}$$

Now, for every i, there is a unique extension $|\cdot|_i$ of $|\cdot|_v$, and each one of these extensions is pairwise non-equivalent. Furthermore, K_i is the completion of L with respect to $|\cdot|_i$ and the embedding φ_i. Finally, for any absolute value $|\cdot|_w$ extending $|\cdot|_v$ to L, there is a unique i such that $|\cdot|_i$ restricted to L is equal to $|\cdot|_w$.

Proof: [1, Proposition 1.3.1.]

□

Corollary 3.3 (Degree formula)**.** If L is a finite dimensional, separable field extension of K, then

$$\sum_{w|v} [L_w : K_v] = [L : K]$$

Proof.
By the primitive element theorem, we can write $L = K(\alpha)$ for some element $\alpha \in L$. Let $P(X)$ be the minimal polynomial of this same α over K. Then, given that L/K is a separable field extension, we can factorize $P(X)$ over K_v like so

$$P(X) = P_1(X) \cdots P_n(X)$$

Now, the embeddings $\sigma : L \hookrightarrow \mathbb{C}_v$ correspond to the maps of α to the roots of $P(X)$, so by proposition 3.1 we have that for each $w \in M_L$ that has been extended from a given $v \in M_K$, $[L_w : K_v] = \text{degree}(P_j(X))$, where j is the unique place such that, borrowing the notation from proposition 3.1, $|\cdot|_w$ is equal to the restriction of $|\cdot|_j$ to L. Therefore

$$[L : K] = deg(P(X)) = \sum_{i=0}^{n} deg(P_i(X)) = \sum_{w|v} [L_w : K_v]$$

□

from this result, we can see why the terms $[L_w : K_v]$ are important for a field extension L/K, and absolute values $w \in M_L$, $v \in K_K$, $w|v$. This motivates the definition of the *local degree of L/K in w*, and the *normalized absolute value associated to v.* They are, respectively

$$n_w = [L_w : K_v] \qquad\qquad \|x\|_w = |x|_w^{n_w}$$

Remark 3.1.2 We will prove Lang's generalization of Roth's theorem, which only deals with finite field extensions of the type $K/\mathbb{Q}$, so we will always have the hypothesis of a finite dimensional, separable field extension. This will make Corollary 3.1.1 useful later.

3.2 The product formula

Definition 3.4. Let L/K be a finite dimensional field extension of degree n. The trace and norm of an element $x \in L$ are defined to be the trace and determinant, respectively, of the endomorphism of the K-vector space L

$$\begin{aligned} T_x : L \to L, \qquad & T_x(\alpha) = x\alpha \\ Tr_{L/K}(x) = Tr(T_x) \qquad & N_{L/K}(x) = \det(T_x). \end{aligned}$$

Note that this definition of $N_{L/K}(x)$ implies $N_{L/K} : L \longrightarrow K$, as L is a K-vector space, hence all the coefficients in the matrix representation of T_x are in K, and therefore $\det(T_x) \in K$. We want to find a formula for $N_{L/K}(x)$ which will be necessary to later prove the product formula. We begin by looking at the characteristic polynomial of T_x

$$f_x(t) = det(tI_d - T_x) = t^n - a_1 t^{n-1} + \cdots + (-1)^n a_n \in K[t]$$

And notice that

$$a_1 = Tr_{L/K}(x) \qquad a_n = N_{L/K}(x)$$

From this observation, we can derive the following proposition.

Proposition 3.5. If L/K is a separable extension and $\sigma : L \to \overline{K}$ varies over the different K-embeddings of L into an algebraic closure $\overline{K}$ of K, then we have, for a given $x \in L$

i) f_x (t) $= \prod_\sigma (t - \sigma(x))$

ii) $Tr_{L/K}(x) = \sum_\sigma \sigma(x)$

iii) $N_{L/K}(x) = \prod_\sigma \sigma(x)$

Proof.
Given L/K is a separable, finite field extension we want to prove that

$$f_x(t) = (p_x(t))^d \qquad d = [L : K(x)] = [L : K]/[K(x) : K]$$

Where $p_x(t)$ is the minimal polynomial of x over K, which we write as

$$p_x(t) = t^m + c_1 t^{m-1} + \cdots + c_m \qquad m = [K(x) : K].$$

Consequently $\{1, x, ..., x^{m-1}\}$ is a basis of $K(x)/K$, and if we call the basis of L/K(x) $\{\alpha_1, ..., \alpha_d\}$, then it is a classic result that the basis of L/K is

$$\alpha_1, x\alpha_1, ..., x^{m-1}\alpha_1; ...; \alpha_d, x\alpha_d, ..., x^{m-1}\alpha_d$$

And in this basis we have

$$T_x \begin{pmatrix} \alpha_1 \\ x\alpha_1 \\ \vdots \\ x^{m-1}\alpha_d \end{pmatrix} = \begin{pmatrix} x\alpha_1 \\ x^2\alpha_1 \\ \vdots \\ (-c_1 x^{m-1} - \cdots - c_m)\alpha_d \end{pmatrix}$$

It is easy to see then that the matrix of $T_x(y) = xy$ with respect to this basis is only made of the same block repeated d times all throughout the diagonal. This block is

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_m & -c_{m-1} & -c_{m-2} & \cdots & -c_1 \end{pmatrix}$$

And one can easily verify it's characteristic polynomial is $p_x(t)$, as it should be, so the characteristic polynomial of the big matrix is $f_x(t) = (p_x(t))^d$ as expected.

Now, L/K is a separable finite field extension, so $L/K(x)$ is also a separable finite field extension, and by the primitive element theorem we can write $L = K(x, y)$ for some $y \in L$. From this, we know that all the K-embeddings of L are uniquely determined by the images of x and y. From the degrees of each field extension, we can then state that the equivalence relation

$$\tau \sim \sigma \iff \tau(x) = \sigma(x)$$

partitions $Hom_K(L, \overline{K})$ into $m = [K(x) : K]$ equivalence classes of $d = [L : K(x)]$ elements each. If we call $\sigma_1, ..., \sigma_m$ the representatives of each equivalence class, we see that

$$p_x(t) = \prod_{i=0}^{m}(t - \sigma_i(x))$$

Which implies

$$f_x(t) = (p_x(t))^d = \prod_{i=0}^{m}(t - \sigma_i(x))^d = \prod_{i=0}^{m}\prod_{\sigma\sim\sigma_i}(t - \sigma(x)) = \prod_{\sigma\in Hom_K(L,\overline{K})}(t - \sigma(x))$$

This proves i), ii), and iii) due to Vietà's equations, which we alluded to before the beginning of the proposition.

□

We follow this by stating without proof an important result on how an absolute value v is extended from a complete field K to a finite dimensional extension K_v, as we will need it to obtain a crucial relationship that we will need for the product formula.

Proposition 3.6. Let K be a complete field relative to an absolute value $|\cdot|_v$ and let L be a finite dimensional field extension of K. Then there is a unique extension of $|\cdot|_v$ to an absolute value $|\cdot|_w$ of L. For any $x \in L$ the equation

$$|x|_w = |N_{L/K}(x)|_v^{1/[L:K]}$$

holds. Moreover, L is complete with respect to $|\cdot|_w$.
Proof.: [1, Proposition1.2.7]

Now, given $v \in M_{\mathbb{Q}}$ we have the following embeddings

$$\begin{array}{c} \mathbb{Q} \hookrightarrow \mathbb{Q}_v \hookrightarrow \overline{\mathbb{Q}_v} \hookrightarrow \mathbb{C}_v \\ |\cdot|_v \mapsto |\cdot|_{v_1} \mapsto |\cdot|_{v_2} \mapsto |\cdot|_{v_3} \end{array}$$

that extend $|\cdot|_v$ to a unique absolute value over $\mathbb{C}_v$, $|\cdot|_{v_3}$. This is because

i) The first embedding is a completion, therefore the absolute value over $\mathbb{Q}_v$ can be defined by $|x|_{v_1} = \lim_{n\to\infty} |x_n|_v$ for a succession $\{x_n\} \subset \mathbb{Q}$, $n \in \mathbb{N}$ with $x = \lim_{n\to\infty} x_n$.

ii) By proposition 3.6, given $\mathbb{Q}_v$ is complete relative to $|\cdot|_{v_1}$.

iii) $\mathbb{C}_v$ is the completion of $\overline{\mathbb{Q}_v}$, so by the same logic as the first embedding, $|\cdot|_{v_3}$ is a unique extension of $|\cdot|_{v_2}$.

Hence $|\cdot|_{v_3}$ is a unique extension of $|\cdot|_v$. Furthermore, it is worth noting that we may assert more generally that an absolute value on a complete field admits a unique extension to its algebraic closure, since the latter is a union of finite dimensional extensions.

$\mathbb{C}_v$ is algebraically closed, therefore every finite field extension of $\mathbb{Q}$ can also be embedded into $\mathbb{C}_v$. Let K be a finite field extension of $\mathbb{Q}$, and let the embeddings of K into $\mathbb{C}_v$ be $\sigma_1, ..., \sigma_N$. Each such embedding can be used to define an absolute value on K, defined by

$$|x|_{\sigma_i} = |\sigma_i(x)|_v$$

which allows us to map the embeddings of K into $\mathbb{C}_v$ to the absolute values:

$$\psi : \{\text{embeddings } K \hookrightarrow \mathbb{C}_v\} \longrightarrow \{w|v \ : w \in M_K,\, v \in M_{\mathbb{Q}}\}.$$

Now, in our proof of Roth's theorem we will only study finite dimensional field extensions of characteristic 0, so by the primitive element theorem we can suppose $K/\mathbb{Q}$ is a finite dimensional field extension and K is generated by a single element ξ. Therefore by proposition 3.2 the map ψ is exhaustive and

$$\#\psi^{-1}(v) = n_v$$

as every absolute value $|\cdot|_w$ extending $|\cdot|_v$ is realized by exactly n_w embeddings
This was the last result we needed in order to prove the product formula, which we will now prove after giving a definition.

Definition 3.7. Given a prime p, and $x \in \mathbb{Q}^*$, let $\operatorname{ord}_p(x)$ be the unique integer such that x can be written as

$$x = p^{ord_p(x)} \cdot \frac{a}{b} \qquad\qquad a, b \in \mathbb{Z},\ \ p \nmid ab$$

If $x = 0$, we set $\operatorname{ord}_p(x) = \infty$ by convention. Then the p-adic absolute value of $x \in \mathbb{Q}$ is

$$|x|_p = p^{-ord_p(x)}$$

Proposition 3.8 (Product formula)**.** Let K be a number field, and let $x \in K^*$. Then

$$\prod_{v\in M_K} \|x\|_v = 1$$

Proof Let $\mathbb{P}$ be the set of all primes, and we begin by showing that the proposition is true for $K = \mathbb{Q}$, as it will be needed afterwards. From the definition of the p-adic absolute value, it is easy to see that

$$\prod_{p\in\mathbb{P}} \|x\|_p = \frac{1}{|x|_\infty}$$

Thus,

$$\prod_{v\in M_{\mathbb{Q}}} \|x\|_v = \prod_{v\in M_{\mathbb{Q}}} |x|_v^1 = |x|_\infty \prod_{p\in\mathbb{P}} |x|_p = 1.$$

We will now prove for a number field K, let $x \in K^*$, therefore $N_{K/\mathbb{Q}}(x) \in \mathbb{Q}^*$ and we have

$$\begin{aligned} 1 &= \prod_{v\in M_{\mathbb{Q}}} |N_{K/\mathbb{Q}}(x)| \\ &= \prod_{v\in M_{\mathbb{Q}}} \prod_{\sigma\in \mathrm{Hom}(K,\mathbb{C}_v)} |\sigma(x)| \quad \text{by proposition 3.5} \\ &= \prod_{v\in M_{\mathbb{Q}}} \prod_{w|v,\, w\in M_K} |x|_w^{n_w} \quad \text{from the exhaustive definition for absolute values} \\ &= \prod_{v\in M_{\mathbb{Q}}} \prod_{w|v,\, w\in M_K} \|x\|_w = \prod_{v\in M_K} \|x\|_w \end{aligned}$$

□

3.3 Heights

Before being able to write the following theorem, we must define the various height functions, they are a way to measure the 'size' of points in projective spaces, or even polynomials.

Given a number field K, for an $\alpha \in K$ we define it's height, relative to K, as the height of the corresponding projective point $(\alpha, 1) \in \mathbb{P}^1(K)$:

$$H_K(\alpha) = \prod_{v\in M_K} \max\{\|\alpha\|_v, 1\}$$

We also define $h_k(\alpha) = \ln H_K(\alpha)$ which often appears in the literature. These definitions clearly depend on the field K, which may not always be desirable, thus, we also have the absolute (multiplicative) height, defined as

$$H(\alpha) = H_K(\alpha)^{1/[K:\mathbb{Q}]}$$

Which can be proven to be independent of the field K using the sum formula. In addition, we also have $h(\alpha) = \ln H(\alpha)$.

Remark 3.3.0. This definition comes from a broader definition for points $P \in \mathbb{P}^n(\mathbb{K})$, which is defined as

$$H_K(P) = \prod_{v\in M_K} \max\{\|x_0\|_v, \|x_1\|_v, ..., \|x_n\|_v\}$$

And is independent of the choice of homogenous coordinates for P, and also $H_K(P) \geq 1$. The second fact comes from the first, and the first comes from the product formula. That said, this height definition is of no relevance to us for Roth's theorem, as we will only look at single elements of number fields, but it gives context to the definition we will use.

Now, given a polynomial $P \in K[X_1, ..., X_m]$ written as

$$P(X_1, ..., X_m) = \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} X_1^{j_1} \cdots X_m^{j_m}$$

We write

$$|P|_v = \max\{|p_{j_1,...,j_m}|_v \ : \ j_h \in \{1,...,r_h\} \ \ \forall h \in 1,...,m\}$$

which in turn allows us to define the same heights for a polynomial $P \in K[X_1, ..., X_m]$:

$$H_K(P) = \prod_{v \in M_K} \max\{1, |P|_v^{n_v}\}$$
$$H(P) = H_K(P)^{/[K:\mathbb{Q}]}$$

For polynomials, too, $h_K(P) = \ln H_K(P)$, $h(P) = \ln H(P)$

Theorem 3.9 (Finiteness of bounded points)**.** For any numbers $A, B \geq 0$, the set

$$\left\{P \in \mathbb{P}^n(\mathbb{Q}) | H(P) \leq A \ \ and \ \ [\mathbb{Q}(P) : \mathbb{Q}] \leq B\right\}$$

is finite. In particular, for any fixed number field K, the set

$$\left\{P \in \mathbb{P}^n(K) | H_K(P) \leq B\right\}$$

is also finite. **Proof.** [2, Theorem B.2.3] □

Proposition 3.10 (Gelfand's inequality)**.** This is a result for projective polynomial heights. Let $r_1, ..., r_m$ be positive integers, and let $P_1, ..., P_s \in \overline{\mathbb{Q}}[X_1, ..., X_m]$ be polynomials with algebraic coefficients such that $deg_{X_i}(P_1 \cdots P_m) \leq r_i \ \ \forall i \in \{1, ..., m\}$. Then

$$\sum_{r=0}^{s} h(P_i) \leq h(P_1 \cdots P_s) + r_1 + \cdots r_m$$

Proof. Gauss' lemma allows us to write, for non-archimedean $v \in S$

$$|P_1 \ldots P_s|_v = |P_1|_v \cdots |P_s|_v$$

On the other hand, for archimedean $v \in S$ we have [1, pages 229-233]

$$\prod_{i=1}^{s} |P_i|_v \leq e^{r_1 + \cdots + r_m} |P_1 \cdots P_s|_v$$

These formulas allow us to compute

$$\begin{aligned}\prod_{i=0}^{s} H_K(P_i) &= \prod_{i=0}^{s} \prod_{v \in M_K} |P_1 \cdots P_s|_v^{n_v} \\ &\leq \prod_{v \in M_K^0} |P_1 \cdots P_s|_v^{n_v} \prod_{v \in M_K^\infty} e^{n_v(r_1 + \cdots + r_m)} |P_1 \cdots P_s|_v^{n_v} \\ &\leq e^{[K:\mathbb{Q}](r_1 + \cdots + r_m)} H_K(P_1 \cdots P_s)\end{aligned}$$

Taking $[K : \mathbb{Q}]^{th}$ roots and then taking logarithms yields Gelfand's inequality. □

Lemma 3.11. Let K/Q be a finite dimensional field extension, and let $x, y \in K$. Then

$$H(x+y) \leq 2H(x)H(y) \quad \text{and} \quad H(xy) \leq H(x)H(y)$$

Proof. We begin with the second inequality:

$$\begin{aligned} H_K(xy) = \prod_{v \in M_K} \max\{1, \|xy\|_v\} &= \prod_{v \in M_K} \max\{1, \|xy\|_v\} \\ &\leq \prod_{v \in M_K} \max\{1, \|x\|_v\} \max\{1, \|y\|_v\} \\ &= H_K(x)H_K(y) \end{aligned}$$

Taking $[K : \mathbb{Q}]$th roots on both sides gives the desired inequality. Now, we begin by defining the following function

$$\varepsilon(v) = \begin{cases} 1 & v \text{ is arquimedean} \\ 0 & v \text{ is not arquimedean} \end{cases}$$

Hence, we write

$$\begin{aligned} H(x+y) &= \prod_{v \in M_K} \max\{1, |x+y|\}_v^{n_v/[K:\mathbb{Q}]} \\ &\leq \prod_{v \in M_K} \max\{1, 2^{\varepsilon(v)}|y|_v, 2^{\varepsilon(v)}|x|_v\}_v^{n_v/[K:\mathbb{Q}]} \\ &\leq \prod_{v \in M_K} 2^{\varepsilon(v)n_v/[K:\mathbb{Q}]} \max\{1, |x|_v\}_v^{n_v/[K:\mathbb{Q}]} \cdot \max\{1, |y|_v\}_v^{n_v/[K:\mathbb{Q}]} \\ &\leq 2H(x)H(y) \end{aligned}$$

□

Lemma 3.12. For every algebraic number α with polynomial $P(x) \in \mathbb{Z}[X]$ there is an integer $M > 0$ such that $M\alpha$ is an algebraic integer.

Proof.

Let $P(x) = \sum_{i=0}^{n} a_i x^i$, making sure $a_n > 0$, and consider $Q(x) = a_n^{n-1}P(x)$. Clearly $Q(\alpha) = 0$, and we can write $Q(\alpha) = \sum_{i=0}^{n} a_i a_n^{n-1-i}(a_n x)^i = 0$, so $a_n\alpha$ is the root of a monic polynomial, with $a_n \in \mathbb{Z}$. □

4 Preliminary results

Throughout the proof of Roth's theorem, previously known lemmas and bounds will be used. To make the actual proof of Roth's theorem read more smoothly, they'll be stated and proven beforehand in this section.

4.1 Notation and defintions

Before moving on to the results however, we will establish the notation we will use hereforward, and give a few definitions that will be used in for the remaining part of the book.
$\lfloor x \rfloor : \mathbb{R} \longrightarrow \mathbb{Z}$ denotes the floor function
$\lceil x \rceil : \mathbb{R} \longrightarrow \mathbb{Z}$ denotes the ceiling function
For a polynomial P, $\deg_{X_h}(P)$ refers to the degree of P in the variable X_h.
Given a polynomial with coefficients in a number field K, $P \in K[X_1, ..., X_m]$ written as

$$P(X_1, ..., X_m) = \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} X_1^{j_1} \cdots X_m^{j_m}$$

We write

$$\partial_{i_1,...,i_m} P = \frac{1}{i_1! i_2! \cdots i_m!} \frac{\partial^{i_1+...+i_m} P}{\partial X_1^{i_1} \cdots X_m^{i_m}}$$

and

$$|P|_v = \max\{|p_{j_1,...,j_m}|_v \ : \ j_h \in \{1, ..., r_h\} \ \ \forall h \in 1, ..., m\}$$

Though we may also write just $|P|$, which is to be understood as $|P|_\infty$.
Let K be a number field, $\alpha = (\alpha_1, ..., \alpha_m) \in K^m$ a point, and $(r_1, ..., r_m) \in \mathbb{Z}^m$ an m-tuple of nonnegative integers. We define the index of a polynomial $P(X_1, ..., X_m) \in K[X_1, ..., X_m]$ with respect to $(\alpha_1, ..., \alpha_m; r_1, ..., r_m)$ as the value

$$\text{Ind}(P) = \min_{i_1,...,i_m} \left\{ \frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m} \mid \partial_{i_1,...,i_m} P(\alpha) \neq 0 \right\}$$

denoted by Ind(P).
δ_{ij} refers to the Kronecker delta that equals 1 iff i=j and is 0 otherwise.

4.2 Results

Lemma 4.1. Given a polynomial $P(X_1, ..., X_m) \in \mathbb{Z}[X_1, ..., X_m]$, and an m-tuple of nonnegative integers $(i_1, ..., i_m)$, the following is true:
a)$\partial_{i_1...i_m} P \in \mathbb{Z}[X_1, ..., X_m]$.
b)If $\deg_{X_h}$(P)$\leq r_h$ for each $1 \leq h \leq m$ then $|\partial_{i_1...i_m} P| \leq 2^{r_1+...+r_m} |P|$.

Proof.
a)We begin by noting that for $i \leq j$:

$$\frac{1}{i!} \frac{d^i X^j}{dX^i} = \binom{j}{i} X^{j-i} \ \text{ defining } \binom{j}{i} = 0 \text{ for } i > j$$

So writing the polynomial P as

$$P(X_1, ..., X_m) = \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} X_1^{j_1} \cdots X_m^{j_m}$$

And differentiating, we obtain the following equation

$$\partial_{i_1,...,i_m} P = \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} X_1^{j_1-i_1} \cdots X_m^{j_m-i_m}$$

Where $p_{j_1,...,j_m} \in \mathbb{Z}$ by hypothesis, and $\binom{j}{i}$ is an integer too, so $\partial_{i_1,..,1_m} P \in \mathbb{Z}[X_1, ..., X_m]$. Now,

$$\binom{j}{i} \leq \sum_{k=0}^{j} \binom{j}{k} = (1+1)^j = 2^j$$

So, from our definition of $|P|$ we have:

$$|\partial_{i_1,...,i_m} P| \leq \max_{j_1,...,j_m,i_1,...,i_m} |p_{j_1,...,j_m}| \cdot \max_{j_1,...,j_m,i_1,...,i_m} \left| \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} \right|$$

$$\leq \max_{j_1,...,j_m,i_1,...,i_m} |p_{j_1,...,j_m}| \cdot \max_{j_1,...,j_m} \left| 2^{j_1+...+j_m} \right| \leq 2^{r_1+...+r_m} |P|$$

As desired.

□

Lemma 4.2. Let $P \in \mathbb{Z}[X_1, ..., X_m]$ with $\deg(X_h) \leq r_h \ \forall h \in \{1, ..., m\}$, and let $\beta = (\beta_1, ..., \beta_m)$ be an m-tuple of algebraic numbers in the number field K. Then, for all m-tuples $j = (j_1, ..., j_m) \in \mathbb{Z}_{\geq 0}^m$ we have

$$H_K(\partial_j P(\beta)) \leq 4^{(r_1+...+r_m)[K:\mathbb{Q}]} H_K(P) \prod_{h=1}^{m} H_K(\beta_h)^{r_h}$$

Proof.

Let $(j_1, ..., j_m) = j$ be any positive m-tuple of integers. Lemma 4.1 allows us to write

$$T(X_1, ..., X_m) = \partial_j P(X_1, ..., X_m) \in \mathbb{Z}[X_1, ..., X_m].$$

and also implies that given another m-tuple of positive integers $(i_1, ..., i_)$ we have

$$|\partial_{i_1,...,i_m} T| = |\partial_{i_1+j_1,...,i_m+j_m} P| \leq 2^{r_1+...+r_m} |P|$$

We now begin by using the triangle inequality for both the archimedean and non-archimedean absolute values. For the archimedean absolute values we only have to consider $v = \infty$

because $T \in \mathbb{Z}[X_1, ..., X_m]$. Now, given $r_h \geq 1$:

$$\begin{aligned}
|T(\beta_1, ..., \beta_m)|_\infty &= \left| \sum_{s_1=1}^{r_1} \cdots \sum_{s_m=1}^{r_m} t_{s_1,...,s_m} \beta_1^{s_1} \cdots \beta_m^{s_m} \right|_\infty \\
&\leq (r_1+1) \cdots (r_m+1) \cdot |T|_\infty \prod_{h=1}^{m} \max\{|\beta_h|_\infty, 1\}^{r_h} \\
&\leq 2^{r_1+...+r_m} \cdot 2^{r_1+...+r_m} |P| \prod_{h=1}^{m} \max\{|\beta_h|_\infty, 1\}^{r_h} \\
&\leq 4^{r_1+...+r_m} \cdot |P| \prod_{h=1}^{m} \max\{|\beta_h|_\infty, 1\}^{r_h}
\end{aligned}$$

For non-archimedean absolute values the inequalities will look different, due to the ultrametric triangle inequality.

$$\begin{aligned}
|T(\beta)|_v &= \left| \sum_{s_1=1}^{r_1} \cdots \sum_{s_m=1}^{r_m} t_{s_1,...,s_m} \beta_1^{s_1} \cdots \beta_m^{s_m} \right|_v \\
&\leq \left| \sum_{s_1=1}^{r_1} \cdots \sum_{s_m=1}^{r_m} \beta_1^{s_1} \cdots \beta_m^{s_m} \right|_v \\
&\leq \prod_{h=1}^{m} \max\{|\beta_h|_v, 1\}^{r_h}
\end{aligned}$$

Every term on the right hand side of the inequality is bigger or equal to 1, which means we can write

$$\max\{|T(\beta_1, ..., \beta_m)|_v, 1\} \leq 4^{r_1+...+r_m} \cdot |P| \prod_{h=1}^{m} \max\{|\beta_h|_v, 1\}^{r_h}$$

For the arquimedian places, and

$$\max\{|T(\beta_1, ..., \beta_m)|_v, 1\} \leq \prod_{h=1}^{m} \max\{|\beta_h|_v, 1\}^{r_h}$$

for the non-archimedean places. Now, for each $v \in M_K$ we raise each side of the inequality to the power of it's corresponding local degree $n_v = [K_v : \mathbb{Q}_v]$ and take the product over all $v \in M_K$ to obtain

$$H_K(T(\beta_1, ..., \beta_m)) \leq 4^{(r_1+...+r_m)[K:\mathbb{Q}]]} \cdot |P|^{[K:\mathbb{Q}]} \prod_{h=1}^{m} H_K(\beta_h)^{r_h}$$

Where we have employed $\sum_{v \in M_K^\infty} n_v = [K : \mathbb{Q}]$ from Corollary 3.1.1. , and the fact that $M_\mathbb{Q}$ only has 1 archimedean place.

□

We note that given P has integer coefficients, $|P|^{[K:\mathbb{Q}]} = H_K(P)$. This is easily verified by noting $\max\{|P|_v, 1\} = 1 \ \forall v \in M_K^0$, and

Lemma 4.3 (Liouville's inequality)**.** Let $K/\mathbb{Q}$ be a number field, and let $\alpha \in K^*$, and let S$\subset M_K$ be a set of absolute values on K^*. Then

$$\prod_{v \in S} \min\{\|\alpha\|_v, 1\} \geq \frac{1}{H_K(\alpha)}$$

Proof.
By the product formula, and using that α is not 0, we have

$$\prod_{v \in M_K} \|\alpha\|_v = 1,$$

From this, and the definition of the height of an element over K we get

$$\begin{aligned} H_K(\alpha) &= \prod_{v \in M_K} \max\{\|\alpha\|_v, 1\} = \prod_{v \in M_K} \|\alpha\|_v \cdot \prod_{v \in M_K} \max\{\frac{1}{\|\alpha\|_v}, 1\} = \prod_{v \in M_K} \max\{\frac{1}{\|\alpha\|_v}, 1\} \\ &= \prod_{v \in M_K} \frac{1}{\min\{\|\alpha\|_v, 1\}} \end{aligned}$$

Taking reciprocals, and from the fact that $S \subset M_K$ we get the result:

$$\frac{1}{H_K(\alpha)} = \prod_{v \in M_K} \min\{\|\alpha\|_v, 1\} \leq \prod_{v \in S} \min\{\|\alpha\|_v, 1\}$$

□

Lemma 4.4. Given an algebraic integer α of degree d over $\mathbb{Q}$ with minimal polynomial

$$Q(X) = X^d + a_{d-1}X^{d-1} + ... + a_1 X + a_0 \in \mathbb{Z}[X]$$

(Which is also the minimal polynomial of α over $\mathbb{Q}$ by Gauss' lemma.) Then for every $r \geq 0$, it is possible to write

$$\alpha^r = a_{d-1}^{(r)}\alpha^{d-1} + a_{d-2}^{(r)}\alpha^{d-2} + ... + a_1^{(r)}\alpha + a_0^{(r)}$$

Where the terms $a_i^{(r)}$ are integers satisfying $|a_i^{(r)}| \leq (|Q| + 1)^r$ for all $i \in \{1, 2, ..., d\}$.
Proof.
The proof will be done by induction on r. Clearly the statement is true for $0 \leq r \leq d-1$ since we can take $a_i^{(r)} = \delta_{ir}$. It is also true for $r = d$ since we can use the minimal polynomial, but it is not necessary for the proof. We assume now that the lemma is true for $r = l$, and will prove it's thus true for $r = l + 1$. By hypothesis:

$$\alpha^{l+1} = \alpha \cdot \sum_{i=0}^{d-1} a_i^{(l)}\alpha^i = a_{d-1}^{(l)}\alpha^d + \sum_{i=0}^{d-2} a_i^{(l)}\alpha^i = a_{d-1}^{(l)} \sum_{i=0}^{d-1} -a_i\alpha^i + \sum_{i=0}^{d-2} a_i^{(l)}\alpha^i$$

Where on the last step we have employed $Q(\alpha) = 0$. Merging the sums we now have:

$$\alpha^{l+1} = \sum_{i=0}^{d-2} \left\{(a_i^{(l)} - a_i a_{d-1}^{(l)})\alpha^i\right\} - a_{d-1}a_{d-1}^{(l)}\alpha^{d-1}$$

Meaning we can set

$$a_i^{(l+1)} = a_i^{(l)} - a_i a_{d-1}^{(l)} \qquad\qquad a_{d-1}^{(l+1)} = -a_{d-1} a_{d-1}^{(l)}$$

And now we're ready to come up with a bound for the coefficients in the case $r = l + 1$:

$$|a_i^{(l+1)}| \leq |a_i^{(l)}| + |a_i| \cdot |a_{d-1}^{(l)}| \leq (|a_i| + 1) \cdot \max\left\{|a_{d-1}^{(l)}|, |a_i^{(l)}|\right\} \leq (|Q| + 1)^l \cdot (|Q| + 1)$$

where the last inequality comes from the induction hypothesis. And so we arrive at the desired bound

$$|a_{d-1}^{(l+1)}| \leq (|Q| + 1)^{l+1}$$

□

Lemma 4.5. Let $P, Q \in K[X_1, ..., X_m]$ be polynomials, and fix both $(r_1, ..., r_m) \in \mathbb{Z}^m$ and a point $(\alpha_1, .., \alpha_m) \in K^m$. The index with respect to $(\alpha_1, .., \alpha_m; r_1, .., r_m)$ has the following properties:

a) $\text{Ind}\ (\partial_{i_1,...,i_m} P) \geq \text{Ind}(P) - \sum_{h=1}^m i_h/r_h$.

b) $\text{Ind}(P + Q) \geq \min\{\text{Ind}(P), \text{Ind}(Q)\}$.

c) $\text{Ind}(PQ) = \text{Ind}\ P + \text{Ind}\ Q$.

Proof.

a) Let $Q = \partial_{i_1,...,i_m} P$, $\alpha = (\alpha_1, ..., \alpha_m)$, and let $(j_1, .., j_m)$ be an m-tuple of integers corresponding to the index of Q with respect to $(\alpha_1, .., \alpha_m; r_1, .., r_m)$, that is, $\text{Ind}(Q) = \dfrac{j_1}{r_1} + ... + \dfrac{j_m}{r_m}$ and $\partial_{j_1,...,j_m} Q(\alpha) \neq 0$. This then implies that $\partial_{i_1+j_1,...,i_m+j_m}(\alpha) \neq 0$ and thus we get $\text{Ind}(P) \leq \text{Ind}(Q) + \sum_{h=1}^m i_h/r_h \Leftrightarrow \text{Ind}(\partial_{i_1,...,i_m} P) \geq \text{Ind}(P) - \sum_{h=1}^m i_h/r_h$

b) Let $(j_1, .., j_m)$ be an m-tuple of integers corresponding to the index of $P + Q$ with respect to $(\alpha_1, .., \alpha_m; r_1, .., r_m)$. This means that at least one of $\partial_{j_1,...,j_m} P(\alpha)$ or $\partial_{j_1,...,j_m} Q(\alpha)$ is non zero, and it must be the one with the smaller index $\text{Ind}(P+Q) = \dfrac{j_1}{r_1} + ... + \dfrac{j_m}{r_m} \geq \min\{\text{Ind}(P), \text{Ind}(Q)\}$

c) We know from the product rule that

$$\partial_{j_1,...,j_m}(PQ) = \sum_{i_1+i'_1} \cdots \sum_{i_m+i'_m} C_{i_1,...,i_m} (\partial_{i_!,...,i_m} P)(\partial_{i'_1,...,i'_m} Q)$$

which implies that given an m-tuple of integers $(j_1, ..., j_m)$ corresponding to the index of PQ, there exists at least one pair of m-tuples $(i_1, ..., i_m)$ and $(i'_1, ..., i'_m)$ such that $i_h + i'_h = j_h\ \ \forall h \in \{1, ..., m\}$ and $(\partial_{i_1,...,i_m} P(\alpha)) \neq 0$, $(\partial_{i'_1,...,i'_m} Q(\alpha)) \neq 0$. This in turn means that the indices of P and Q with respect to $(\alpha_1, .., \alpha_m; r_1, .., r_m)$ will satisfy $\text{Ind}(P) \leq \dfrac{i_1}{r_1} + ... + \dfrac{i_m}{r_m}$ and $\text{Ind}\ Q \leq \dfrac{i'_1}{r_1} + ... + \dfrac{i'_m}{r_m}$, this gives the first of the two inequalities: Ind $PQ \geq \text{Ind}(P) + \text{Ind}(Q)$.

To obtain the opposite inequality, we begin by looking at the m-tuples for P and Q. The set of m-tuples corresponding to the index of P is finite, so it has a minimum, we will call the minimum of this set ordered lexicographically $(l_1, ..., l_m)$. This means that for any other m-tuple $(i_1, ..., i_m)$ in the set, there exists an $h \in \{1, ..., m\}$ such that $i_h > l_h$. We do the same for Q, and call the minimum from it's set of m-tuples corresponding to Ind Q $(l'_1, ..., l'_m)$. We now define $(j_1, ..., j_m) = (l_1 + l'_1, ..., l_m + l'_m)$ and notice that due to taking the minimum we have

$$\begin{aligned}\partial_{j_1,...,j_m}(PQ)(\alpha) &= \sum_{i_1+i'_1} \cdots \sum_{i_m+i'_m} C_{i_1,...,i_m}(\partial_{i_1,...,i_m}P(\alpha))(\partial_{i'_1,...,i'_m}Q(\alpha)) \\ &= (\partial_{l_1,...,l_m}P(\alpha))(\partial_{l'_1,...,l'_m}Q(\alpha)) \neq 0\end{aligned}$$

Which gives an upper bound for the index of PQ: $\text{Ind}(PQ) \leq \frac{j_1}{r_1} + \cdots + \frac{j_m}{r_m} = \frac{l_1 + l'_1}{r_1} \cdots \frac{l_m + l'_m}{r_m} = \text{Ind}(P) + \text{Ind}(Q)$.

Combining both inequalities gives the desired result

□

As an aside, the index of the 0 polynomial is infinity, due to never being non zero regardless of how many times it is differentiated, which also means only 0 has Index infinity. It is worth noting that with this, we have proven that the index is a valuation from the polynomials of m variables to $\mathbb{Q}$. This is important, because this proof of Roth's theorem hinges on reaching a contradiction through the use of the index of a certain polynomial, and that can be done because the index has these 'nice' properties.

Lemma 4.6 (Siegel's Lemma)**.** Let A = (a_{ij}) be a matrix with coefficients in $\mathbb{Z}$, whose absolute value is defined to be $|A| = \max_{j,k}\{|a_{jk}|\}$. Let z = $(z_1, ..., z_N)$ and we similarly define it's absolute value to be $|z| = max\,(|z_1|, ..., |z_N|)$.

The lemma states that given an M × N matrix A with coefficients in $\mathbb{Z}$, not all 0, and assuming N > M, then there exists a vector Z $\in \mathbb{Z}^N$ satisfying

$$Az = 0, \qquad z \neq 0, \qquad |z| \leq (N|A|)^{\frac{M}{N-M}}$$

Proof. Let Z $= (N|A|)^{\frac{M}{N-M}}$, and $z = (z_1, ..., z_N) \in \mathbb{Z}^N$ be any vector such that $0 \leq z_i \leq Z$ $\forall i \in \{1, ..., N\}$. Given $|a_{jk}| \leq |A|, 0 \leq z_i \leq Z$, the matrix is M x N, and accounting for zeroes, Az takes at most $(N|A|Z+1)^M$ distinct values. Now, due to the choice of Z we have $(N|A|Z+1)^M \leq (N|A|)^M(Z+1)^M = Z^{N-M}(Z+1)^M < (Z+1)^N$, meaning the cardinality of the set of distinct possible vectors z is strictly bigger than the cardinality of the set of distinct values Az can take. Hence, there exist $z^{(1)} \neq z^{(2)}$ such that $Az^{(1)} = Az^{(2)}$, and $z := z^{(1)} - z^{(2)}$ satisfies the conditions of the lemma.

□

Lemma 4.7. Given $r_1, ..., r_m \in \mathbb{Z}$, and a fixed $\varepsilon \in (0,1)$, then there are at most

$$(r_1 + 1) \cdots (r_m + 1) \cdot e^{-m\varepsilon^2/4}$$

m-tuples of integers $(i_1, ..., i_m)$ that satisfy the following conditions

$$0 \leq i_j \leq r_j \quad \forall j \in \{1, 2, ..., m\} \qquad\qquad \frac{1}{m}\sum_{j=0}^{m} \frac{i_j}{r_j} \leq \frac{1}{2} - \varepsilon$$

Proof.

Let $I(m, \varepsilon)$ denote the set of m-tuples we want to count. Clearly then we have

$$\#I(m, \varepsilon) = \sum_{(i_1,...,i_m)\in I(m,\varepsilon)} 1 \qquad \leq \sum_{(i_1,...,i_m)\in I(m,\varepsilon)} \exp \frac{\varepsilon}{2}\left(\frac{m}{2} - m\varepsilon - \sum_{j=0}^{m} \frac{i_j}{r_j}\right)$$

Where we use the hypothesis that $0 \leq \frac{m}{2} - m\varepsilon - \sum_{j=0}^{m} \frac{i_j}{r_j}$ and $e^t \geq 1$ for $t \geq 0$. We continue, expanding now the sum to all possible i-tuples

$$\leq \sum_{i_1=0}^{r_1} \cdots \sum_{i_m=0}^{r_m} \exp \frac{\varepsilon}{2}\left(\frac{m}{2} - m\varepsilon - \sum_{j=0}^{m} \frac{i_j}{r_j}\right) = exp\left(\frac{-m\varepsilon^2}{2}\right) \sum_{i_1=0}^{r_1} \cdots \sum_{i_m=0}^{r_m} \exp \frac{\varepsilon}{2}\left(\frac{m}{2} - \sum_{j=0}^{m} \frac{i_j}{r_j}\right)$$

And using the very fact that we're adding over every possible configuration of $(i_1, ..., i_m)$, the latest sum is equal to

$$exp\left(\frac{-m\varepsilon^2}{2}\right) \prod_{h=1}^{m} \left(\sum_{i_m=0}^{r_h} \exp \frac{\varepsilon}{2}\left(\frac{1}{2} - \frac{i}{r_h}\right)\right)$$

Now, given $e^t \leq 1 + t + t^2$ for all $|t| \leq 1$ we can find a bound for the summation above

$$\begin{aligned}
\sum_{i=0}^{r} \exp \frac{\varepsilon}{2}\left(\frac{1}{2} - \frac{i}{r}\right) &\leq \sum_{i=0}^{r} \left\{\left(1 + \frac{\varepsilon}{4} + \frac{\varepsilon^2}{16}\right) - \left(\frac{\varepsilon}{2} + \frac{\varepsilon^2}{4}\right)\frac{i}{r} + \frac{\varepsilon^2 i^2}{4r^2}\right\} \\
&\leq (r+1)\left(1 + \frac{\varepsilon^2}{48} + \frac{\varepsilon^2}{12r}\right) \\
&\leq (r+1)\left(1 + \frac{\varepsilon^2}{4}\right)
\end{aligned}$$

Substituting back into the product, we get

$$\begin{aligned}
\#I(m, \varepsilon) &\leq exp\left(\frac{-m\varepsilon^2}{2}\right) \prod_{h=1}^{m} \left((r_h + 1)\left(1 + \frac{\varepsilon^2}{4}\right)\right) \\
&\leq exp\left(\frac{-m\varepsilon^2}{2}\right) \prod_{h=1}^{m} \left((r_h + 1) \exp \frac{\varepsilon^2}{4}\right) \\
&= (r_1 + 1) \cdots (r_m + 1) \exp\left(\frac{-m\varepsilon^2}{4}\right)
\end{aligned}$$

□

Before finally embarking on the endeavour of proving Roth's theorem, we will prove two lemmas to simplify our task. The first is the reduction lemma, which reduces the scope of the proof to the algebraic integers, and the second one is a result equivalent to Roth's theorem, which is the one we will prove. It is worth mentioning beforehand that for this generalization of the original version of Roth's theorem, we specifically consider the field extension $K/\mathbb{Q}$, so every algebraic number we consider has a minimal polynomial in $\mathbb{Z}[X]$.

Lemma 4.8. If Roth's theorem is true for all algebraic integers, then it is true for all algebraic numbers.

Proof.

We will prove this by contradiction. Let $\alpha \in K$ be an algebraic number, and suppose Roth's theorem is false for α, this implies there exist infinitely many $\beta \in K$ satisfying (1). Now, for a given $\beta \in K$ that satisfies (1), we affirm that

$$\infty = \#\big\{\beta \in K \mid \beta \text{ satisfies } (1.1)\big\} \leq \sum_{S' \subset S} \#\left\{\beta \in K \mid \prod_{v \in S'} \|\beta - \alpha\|_v \leq \frac{1}{H_K(\beta)^{2+\varepsilon}}\right\} \quad (4.1)$$

Because for every $\beta \in K$ that satisfies (1), there is always some non-empty subset of S, $S_\beta := \big\{v \in S : \|\beta - \alpha\|_v \leq 1\big\}$ such that

$$\prod_{v \in S} \min\big\{1, \|\beta - \alpha\|_v\big\} = \prod_{v \in S_\beta} \|\beta - \alpha\|_v \leq \frac{1}{H_K(\beta)^{2+\varepsilon}}$$

By Hypothesis of Roth's Thorem, $S \subset M_K$ is finite, however the sum in (4.1) is infinite, so by the pidgeon hole principle there's at least one of the terms in the sum that's infinite. From this, we deduce that for this lemma, after replacing S by one of it's subsets (which will henceforth be called S), we can further assume there are infinitely many $\beta \in K$ such that, for every $\varepsilon > 0$

$$\prod_{v \in S} \|\beta - \alpha\|_v \leq \frac{1}{H_K(\beta)^{2+\varepsilon}}$$

Now we choose an integer $D > 0$ such that $D\alpha$ is an algebraic integer, and let $\beta \in K$ satisfy both (1) and $H_K(\beta) > H_K(D)^{1+6}$. Finally, From the definition of $H_K(\cdot)$ we get $H_K(D\beta) \leq H_K(D)H_K(\beta)$ and

$$\prod_{v \in S} \|D\|_v \leq \prod_{v \in S} \max\{\|D\|_v, 1\} \leq \prod_{v \in M_K} \max\{\|D\|_v, 1\} = H_K(D)$$

With all of this, we can finally show that Roth's theorem is false for $D\alpha$, finishing the proof.

$$\begin{aligned}\prod_{v \in S} \|D\beta - D\alpha\|_v &\leq \frac{H_K(D)}{H_K(\beta)^{2+\varepsilon}} = \frac{H_K(D)}{H_K(\beta)^{2+\varepsilon/2}} \cdot \frac{1}{H_K(\beta)^{\varepsilon/2}} \\ &\leq \frac{H_K(D)}{(H_K(D\beta)/H_K(D))^{2+\varepsilon}} \cdot \frac{1}{(H_K(D)^{1+6/\varepsilon})^{\varepsilon/2}} = \frac{1}{H_K(D\beta)^{2+\varepsilon/2}}\end{aligned}$$

Hence, if Roth's theorem is true for algebraic integers, then it is also true for algebraic numbers.

□

Theorem 4.9. *Let K be a number field, $S \subset M_K$ a finite set of absolute values over K, with each one of them extended in some way to $\overline{K}$. Given $\alpha \in \overline{K}$, $\varepsilon > 0$, and a function $\zeta : S \to [0,1]$ such that*

$$\sum_{v \in S} \zeta_v = 1$$

Then there are finitely many $\beta \in K$ such that

$$\|\beta - \alpha\|_v \leq \frac{1}{H_K(\beta)^{(2+\varepsilon)\zeta_v}} \quad \forall v \in S \tag{2}$$

We will see shortly that this Theorem is in fact equivalent to Roth's theorem.

Theorem 4.10. *Theorem 2.3 is true if and only if Theorem 4.9 is true.*
Proof We will first assume Theorem 2.1 to be true, and then assume true the hypotheses of Theorem 4.7. We denote by $\zeta : S \to [0,1]$ the function from the hypotheses of theorem 4.7, and then assume $\beta \in K$ satisfies (2). Multiplying $\|\beta - \alpha\|$ over $v \in S$ and using $\sum_{v \in S} \zeta_v = 1$ we get

$$\prod_{v \in S} \min\{1, \|\beta - \alpha\|_v\} \leq \prod_{v \in S} \|\beta - \alpha\|_v \leq \frac{1}{H_K(\beta)^{2+\varepsilon}}$$

And applying Theorem 2.1 we get that there are only finitely such β's in K that satisfy (2).

We now assume Theorem 4.7 is true, and that there are infinitely many $\beta \in K$ satisfying (1). We will prove the assertion by contradiction, by showing that each one of those β's must satisfy (2) for at least one $v \in S$, meaning infinite $\beta \in K$ satisfy (2), contrary to our hypothesis that Theorem 4.7 is true. Let $s = \#S$, and consider

$$\zeta_v : S \to [0,1], \;\; \zeta_v = \frac{a_v}{s} \text{ with } a_v \in \mathbb{Z}_{\geq 0} \text{ and } \sum_{v \in S} a_v = s$$

Clearly there's a finite amount of these maps, which will be denoted by $\mathscr{Z}$.Now let $\beta \in K$ satisfy (1), we want to show this implies it satisfies (2) for one of the maps in $\mathscr{Z}$. For every $v \in S$ we define a real number $\lambda_v(\beta)$ by the formula

$$\min\{1, \|\beta - \alpha\|_v\} = \frac{1}{H_K(\beta)^{(2+\varepsilon)\lambda_v(\beta)}}$$

which ensures that $\lambda_v(\beta) \geq 0$. We now multiply over S, keeping in mind we have assumed Roth's theorem to be true

$$\prod_{v \in S} \min\{1, \|\beta - \alpha\|_v\} = \frac{1}{H_K(\beta)^{(2+\varepsilon)\sum_v \lambda_v(\beta)}} \leq \frac{1}{H_K(\beta)^{(2+\varepsilon)}}$$

And we obtain that $\sum_v \lambda_v(\beta) \geq 1$, from this

$$\sum_{v \in S} 2s\lambda_v(\beta) \geq 2s \geq s$$

This implies that we can always find integers $a_v(\beta)$ satisfying our desired properties from before:

$$0 \leq a_v(\beta) \leq 2s\lambda_v(\beta) \text{and} \sum_{v \in S} a_v(\beta) = s$$

Meaning $a_v(\beta)/s$ belongs to $\mathcal{Z}$, as desired. Therefore, the infinite amount of β's in K that satisfy (2) also satisfy (1) for some $\varphi_v \in \mathcal{Z}$, contradiction. This completes the proof of the theorem. □

5 Proof of Roth's Theorem

5.1 Construction of the Auxiliary Polynomial

We begin the proof by constructing a Polynomial $P(X_1, ..., X_m) \in \mathbb{Z}[X_1, ..., X_m]$ with bounded coefficients that vanishes to high order at $(\alpha, ..., \alpha) \in K^n$. We note that this is the equivalent of step 1 from Liouville, 1844.

Proposition 5.1. Let α be an algebraic integer of degree d over $\mathbb{Q}$, let $\varepsilon > 0$ be a fixed constant, and let m be an integer such that

$$e^{m\varepsilon^2/4} > 2d$$

Let also $r_1, ..., r_m$ be given positive integers. Then there exists a polynomial $P(X_1, ..., X_m) \in \mathbb{Z}[X_1, ..., X_m]$ satisfying the following conditions

i) P has degree at most r_h in the variable X_h $\forall h \in \{1, ..., m\}$

ii) The index of P with respect to $(\alpha, ..., \alpha; r_1, ..., r_m)$ satisfies

$$\text{Ind}(P) \geq \frac{m}{2} \cdot (1 - 2\varepsilon)$$

iii)

$$|P|_\infty \leq B(\alpha)^{r_1 + \cdots + r_m}$$

Where $B(\alpha)$ is a constant that only depends on α

Proof.
As usual we will write

$$P(X_1, ..., X_m) = \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} X_1^{j_1} \cdots X_m^{j_m}$$

Where this time the coefficients $P_{j_1,...,j_m}$ are unknown and must be determined. Accounting for any coefficients that may be 0, the number of coefficients is

$$N = (r_1 + 1) \cdots (r_m + 1)$$

As shown in lemma 4.1, differentiation with respect to an m-tuple $(i_1, .., i_m)$ yields

$$\partial_{i_1,...,i_m} P = P_{i_1,...,i_m} = \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} X_1^{j_1 - i_1} \cdots X_m^{j_m - i_m}$$

Which we now evaluate at $(\alpha, ..., \alpha)$ to have the polynomial equal 0 and use that to determine it's coefficients. Using now lemma 4.4 to express the powers of α larger than d-1 gives

$$\begin{aligned}
&P_{i_1,...,i_m}(\alpha, ..., \alpha) \\
&= \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} \alpha^{j_1 - i_1 + \cdots + j_m - i_m} \\
&= \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} \left(\sum_{k=1}^{d} a_k^{(j_1 + \cdots + j_m - i_1 - \cdots - i_m)} \alpha^k \right) \\
&= \sum_{k=1}^{d} \left(\sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} a_k^{(j_1 + \cdots + j_m - i_1 - \cdots - i_m)} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} \right) \alpha^k
\end{aligned}$$

Thus, $P_{i_1,...,i_m}(\alpha,...,\alpha) = 0$ if the sums within the big parenthesis from the previous expression equal 0. This is equivalent to saying $P_{i_1,...,i_m}(\alpha,...,\alpha) = 0$ if we choose $p_{j_1,...,j_m}$ to satisfy the linear system of equations

$$\sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} p_{j_1,...,j_m} a_k^{(j_1+\cdots+j_m-i_1-\cdots-i_m)} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} = 0$$

For every $k \in \{1,...,d\}$.
We are now in a spot to impose the conditions we desire from our polynomial, and see what coefficients $p_{j_1,...,j_m}$ the linear system of equations yields. To obtain a polynomial satisfying ii), we need $P_{i_1,...,i_m}(\alpha,...,\alpha) = 0$ for all m-tuples satisfying

$$\frac{i_1}{r_1} + \cdots \frac{i_m}{r_m} \leq \frac{m}{2} \cdot (1-2\varepsilon) = \frac{m}{2} - m\varepsilon.$$

Lemma 4.7 tells us that there are at most $(r_1+1)\cdots(r_m+1)e^{-m\varepsilon^2/4}$ such m-tuples, therefore, to obtain a Polynomial that satisfies ii), it suffices to choose coefficients $p_{j_1,...,j_m}$ that satisfy a system of M linear equations with integer coefficients

$$M \leq d \cdot (r_1+1)\cdots(r_m+1)e^{-m\varepsilon^2/4} = dNe^{-m\varepsilon^2/4} \leq \frac{dN}{2d} = \frac{N}{2} < N$$

Where M is bounded by the possibility of having d linear equations for each possible m-tuple. We have $M < N$, so all that is left is to find a bound for the coefficients of the linear equations, and we can use Siegel's lemma to obtain the coefficients of the polynomial with our desired properties.

By lemma 4.4 again, we know that $|a_k^{(l)}| \leq (|Q|+1)^l$, where Q is the minimal polynomial of α over $\mathbb{Q}$, so using the same bound as before for the binomial coefficients we get

$$\begin{aligned} \left| \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} a_k^{(j_1+\cdots+j_m-i_1-\cdots-i_m)} \right| &\leq 2^{j_1+\cdots+j_m}(|Q|+1)^{(j_1+\cdots+j_m)} \\ &\leq (2|Q|+2)^{r_1+\cdots+r_m} \end{aligned}$$

Applying now Siegel's lemma yields the existence of a polynomial P such that

$$\begin{aligned} |P|_\infty &\leq \left(N(2|Q|+2)^{r_1+\cdots+r_m}\right)^{M/N-M} \\ &\leq N(2|Q|+2)^{r_1+\cdots+r_m} \\ &\leq 2^{r_1+\cdots+r_m}(2|Q|+2)^{r_1+\cdots+r_m} \\ &= B(\alpha)^{r_1+\cdots+r_m} \end{aligned}$$

Where $B(\alpha)$ is a constant that depends only on α, and with this P satisfies conditions i), ii), and iii).

□

5.2 The index of the Auxiliary Polynomial must be large

After proving the existence of a Polynomial with desirable properties, we now prove that if $(\beta_1,...,\beta_m) \in K^n$ are 'close' to $(\alpha,...,\alpha) \in \overline{K}^n$, then the polynomial will vanish to high order at $(\beta_1,...,\beta_m)$, that is, Ind(P) with respect to $(\beta_1,...,\beta_m; r_1,...,r_m)$ is large. Notice how this is the equivalent of step 2 from *Liouville*, 1844.

Proposition 5.2 (Large index)**.** Given $0 < \delta < 1$, and given ε such that $0 < \varepsilon < \delta/28$. Let $\alpha \in \overline{K}$ be an algebraic integer over $\mathbb{Q}$ and let m be an integer satisfying once again $e^{m\varepsilon^2/4} > 2d$. Let also $r_1, ..., r_m$ be given positive integers, and use proposition 5.1 to choose a polynomial $P(X_1, ..., X_m) \in \mathbb{Z}[X_1, ..., X_m]$ satisfying the properties given in that same proposition. With that done, let $S \subset M_K$ be a finite set of absolute values on K, with each one of them extended in some way to $\overline{K}$, and let

$$\zeta : S \to [0, 1] \quad \text{satisfy} \quad \sum_{v \in S} \zeta_v = 1 \tag{5.2-i}$$

Suppose that $\beta_1, ..., \beta_m \in K$ satisfy

$$\|\beta_h - \alpha\|_v \leq \frac{1}{H_K(\beta_h)^{(2+\delta)\zeta_v}} \ \forall v \in S, \ \forall h \in 1, .., m \tag{5.2.-ii}$$

and suppose further that

$$\max_{1 \leq h \leq m} \{H_K(\beta_h)^{r_h}\} \leq \min_{1 \leq h \leq m} \{H_K(\beta_h)^{r_h}\}^{1+\varepsilon} \tag{5.2-iii}$$

and tha there exists a constant $C(\alpha, \delta)$ such that

$$C(\alpha, \delta) \leq H(\beta_h) \ \forall h \in 1, ..., m \tag{5.2-iv}$$

If all of these hypotheses are met, then the index of P with respect to $(\beta_1, ..., \beta_m; r_1, ..., r_m)$ is large:

$$\text{Ind}(P) \geq \varepsilon m$$

Recall from Theorems 4.10 and 4.11, that the hypotheses before and including 5.2-i imply a finite number of $\beta \in K$ can satisfy 5.2-ii. Assuming both true later allows us to seamlessly use this proposition when we assume there are infinite $\beta \in K$ that satisfy (1), and this proposition is key to finding a contradiction from that hypothesis. It is worth remarking here too, that the structure of this proof of Roth's theorem implies choosing the β_i that approximate α 'well enough', meaning we choose them before the positive integers $r_1, ..., r_m$, so we must be vigilant regarding the dependence on these integers. Now, before moving on to the proof of this proposition, we need one more lemma.

Lemma 5.3. Let $r_1, ..., r_m$ be given positive integers, let $P(X_1, ..., X_m) \in \mathbb{Z}[X_1, ..., X_m]$ be a polynomial such that $deg_{X_h}(P) \leq r_h$, and let $\theta = \text{Ind}(P)$ with respect to $(\alpha, ..., \alpha; r_1, ..., r_m)$. Let $0 < \delta < 1$ be another given constant and choose θ_0 such that $0 < \theta_0 < \theta$. Let $S \subset M_K$ be a finite subset of absolute values on K with each one extended in some way to $\overline{K}$, and suppose both hypotheses 5.2-i and 5.2-ii are satisfied. We will write D = $\min_h\{H_K(\beta_h)^{r_h}\}$, and let $j = (j_1, ..., j_m)$ be an m-tuple satisfying

$$\sum_{h=1}^{m} \frac{j_h}{r_h} \leq \theta_0$$

With all this, then

$$\prod_{v \in S} \|\partial_j P(\beta_1, ..., \beta_m)\|_v \leq (4H(\alpha))^{[K:\mathbb{Q}](r_1 + \cdots + r_m)} H_K(P) D^{-(2+\delta)(\theta - \theta_0)}$$

Proof. Let T = $\partial_j P$. We want to apply Taylor's theorem around $(\alpha, ..., \alpha)$ in a step reminiscent of Liouville, 1844. However, we need bounds for the Taylor coefficients and absolute values $v \in M_K$ defined at $\alpha \in \overline{K}$. For this reason, we will henceforth suppose every

$v \in M_K$ has been extended in some way to $K(\alpha)$ in order to estimate $|\partial_{i_1,...,i_m} T(\alpha,...,\alpha)|_v$ for the considered m-tuples $(i_1,...,i_m)$. We note that $|\partial_{i_1,...,i_m} T(\alpha,...,\alpha)|_v$ is the sum of at most $(r_1+1)\cdots(r_m+1)$ terms, each of which has a magnitude satisfying

$$\begin{aligned} |\partial_{i_1,...,i_m} T|_\infty \max\{1,|\alpha|_v\}^{r_1+\cdots+r_m} &= |\partial_{i_1+j_1,...,i_m+j_m} P|_\infty \max\{1,|\alpha|_v\}^{r_1+\cdots+r_m} \\ &\le |P|_\infty (2\max\{1,|\alpha|_v\})^{r_1+\cdots+r_m} \end{aligned}$$

Now, we recall that for $r \le 1$, $(r+1) \le 2^r$, so

$$\begin{aligned} |\partial_{i_1,...,i_m} T(\alpha,...,\alpha)|_v &\le (r_1+1)\cdots(r_m+1)|P|_\infty (2\max\{1,|\alpha|_v\})^{r_1+\cdots+r_m} \\ &\le |P|_\infty (4\max\{1,|\alpha|_v\})^{r_1+\cdots+r_m} \end{aligned}$$

Using this bound, we will confirm that just as P is a polynomial that vanishes to high order at $(\alpha,...,\alpha)$, T also vanishes to a fairly high order at $(\alpha,...,\alpha)$. From lemma 4.5 we have the following inequality for the index of T with respect to $(\alpha,...,\alpha;r_1,...,r_m)$

$$\mathrm{Ind}(T) = \mathrm{Ind}(\partial_j P) \ge \mathrm{Ind}(P) - \sum_{h=1}^{m} \frac{j_h}{r_h} \ge \theta - \theta_0$$

The Taylor expansion of T about $(\alpha,...,\alpha)$ is thus

$$T(X_1,...,X_m) = \sum_{\substack{i_1=0 \\ \frac{i_1}{r_1}+\cdots+\frac{i_m}{r_m} \ge \theta-\theta_0}}^{r_1} \cdots \sum_{i_m}^{r_m} \partial_{i_1,...,i_m} T(\alpha,...\alpha)(X_1-\alpha)^{i_1}\cdots(X_m-\alpha)^{i_m}$$

Which we evaluate at $X_h = \beta_h$ for each absolute value $v \in S$, with $n_v = [K_v : \mathbb{Q}_v]$

$$\begin{aligned} &T(\beta_1,...,\beta_m) \\ &\quad\le \sum_{\substack{i_1=0 \\ \frac{i_1}{r_1}+\cdots+\frac{i_m}{r_m} \ge \theta-\theta_0}}^{r_1} \cdots \sum_{i_m}^{r_m} |\partial_{i_1,...,i_m} T(\alpha,...\alpha)|_v |(\beta_1-\alpha)|_v^{i_1}\cdots|\beta_m-\alpha|_v^{i_m} \\ &\quad\le (r_1+1)\cdots(r_m+1) \max_{\frac{i_1}{r_1}+\cdots+\frac{i_m}{r_m} \ge \theta-\theta_0} \{|\partial_{i_1,...,i_m} T(\alpha,...\alpha)|_v |(\beta_1-\alpha)|_v^{i_1}\cdots|\beta_m-\alpha|_v^{i_m}\} \\ &\quad\le 4^{r_1+\cdots+r_m}|P|_v \max_{\frac{i_1}{r_1}+\cdots+\frac{i_m}{r_m} \ge \theta-\theta_0} \left\{ \frac{1}{\left(H_K(\beta_1)^{i_1/n_v}\cdots H_K(\beta_m)^{i_m/n_v}\right)^{(2+\delta)\zeta_v}} \right\} \end{aligned}$$

We now bound the product of heights to make this bound easier to use

$$H_K(\beta_1)^{i_1/n_v}\cdots H_K(\beta_m)^{i_m/n_v} = \left(H_K(\beta_1)^{r_1}\right)^{i_1/r_1 n_v}\cdots\left(H_K(\beta_m)^{r_m}\right)^{i_m/r_m n_v} \ge D^{(\theta-\theta_0)/n_v}$$

From this we get

$$|T(\beta_1,...,\beta_m)|_v \le \frac{4^{r_1+\cdots+r_m}|P|_v}{D^{(\theta-\theta_0)(2+\delta)\zeta_v/n_v}}$$

We can raise each side of this inequality to the power n_v, multiply over all $v \in S$, and using that $\sum_{v\in S}\zeta_v = 1$ we finally obtain the desired bound

$$\prod_{v\in S}\|T(\beta_1,...,\beta_m)\|_v \le \frac{4^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}H_K(P)}{D^{(\theta-\theta_0)(2+\delta)}}$$

□

We can now move on to proving Proposition 5.2.

Proof. Let j $= (j_1, ..., j_m)$ be m-tuple such that $\sum_{h=1}^m j_h/r_h \leq m\varepsilon$. We will show that $\partial_j P(\beta_1, ..., \beta_m) = 0$ (The index w.r.t $(\beta_1, ..., \beta_m; r_1, ..., r_m)$ is also large). We must note first that $0 < \delta < 1$ and $0 < \varepsilon < \delta/28 \implies 0 < \varepsilon < 1/4 \implies \theta = m(1-2\varepsilon)/2 > m\varepsilon = \theta_0$, so we can use lemma 5.3 for the m-tuple we have chosen. Lemma 5.3 gives, recalling the definition of T:

$$\begin{aligned}\prod_{v\in S} \|\partial_j(\beta_1, ..., \beta_m)\|_v &\leq \frac{4^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}H_K(P)}{D^{(\theta-\theta_0)(2+\delta)}} \\ &\leq \frac{(4B(\alpha))^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}}{D^{(m(2+\delta)(1-2\varepsilon)/2-m\varepsilon)}}\end{aligned}$$

Where we have used the properties ii) and iii) of the auxiliary polynomial. On the other hand lemma 4.2 says

$$\begin{aligned}H_K(\partial_j P(\beta_1, ..., \beta_m)) &\leq 4^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}H_K(P)\prod_{h=1}^m H_K(\beta_h)^{r_h} \\ &\leq (4B(\alpha))^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}D^{m(1+\varepsilon)}\end{aligned}$$

Where we also used one of the hypotheses of this proposition in the last inequality. Now, lemma 4.3, due to it's own hypotheses tells us that either $\partial_j P(\beta_1, ..., \beta_m) = 0$, or

$$\prod_{v\in S} \|\partial_j P(\beta_1, ..., \beta_m)\|_v \geq H_K(\partial_j P(\beta_1, ..., \beta_m))^{-1}$$

We want to show that $\partial_j P(\beta_1, ..., \beta_m) = 0$, so we will assume the inequality and reach a contradiction. Liouville's inequality provides a link between the two inequalities we had derived, and implies

$$D^{m(2+\delta)(0.5-2\varepsilon)-(1+\varepsilon)} \leq (4B(\alpha))^{2(r_1+\cdots+r_m)[K:\mathbb{Q}]}$$

From $0 < \delta < 1$ and $0 < \varepsilon < \delta/28$ we get

$$(0.5 - 2\varepsilon)(2 + \delta) - (1 + \varepsilon) > \delta/2 - 5\varepsilon - 2\varepsilon\delta > \delta/2 - 7\varepsilon > \delta/2 - \delta/4 = \delta/4$$

Which implies

$$D^{m\delta/4} < (4B(\alpha))^{2(r_1+\cdots+r_m)[K:\mathbb{Q}]}$$

and, together with another hypothesis from this proposition

$$\max_{1\leq j\leq m}\{H_K(\beta_h)^{r_h}\} \leq D^{1+\varepsilon} < (4B(\alpha))^{8(r_1+\cdots+r_m)[K:\mathbb{Q}](1+\varepsilon)/m\delta}.$$

Choosing now j such that $r_j = \max_{h=1}^m\{r_h\}$ gives us the following inequality:

$$H_K(\beta_j) < (4B(\alpha))^{8[K:\mathbb{Q}](1+\varepsilon)/\delta}$$

Which allows us to choose a constant $C(\alpha, \delta)$ to complete the proof. Recall that hypothesis 5.2-iv is to assume that for every $h \in \{1, ..., m\}$ $C(\alpha, \delta) \leq H_K(\beta_h)$. since $H_K(\beta_j) < (4B(\alpha))^{8[K:\mathbb{Q}](1+\varepsilon)/\delta}$ choosing $C(\alpha, \delta)$ sufficiently large we obtain the desired contradiction, which implies that $\partial_j P(\beta_1, ..., \beta_m) = 0$.

If we wanted an explicit $C(\alpha, \delta)$, the following would suffice

$$C(\alpha, \delta) = \min_{\delta}\{(4B(\alpha))^{8[K:\mathbb{Q}](1+\varepsilon)/\delta}\} = (4B(\alpha))^8$$

Where we also considered $[K : \mathbb{Q}] \geq 1$.

□

5.3 The index of the Auxiliary Polynomial must be small

We just showed that our auxiliary polynomial which vanishes to high order at $(\alpha, ..., \alpha)$, also vanishes to fairly high order at $(\beta_1, ..., \beta_m)$. We will now show that the opposite must be true given our hypotheses, and this contradiction will be the cornerstone of our proof that that there aren't an infinite number of points $(\beta_1, ..., \beta_m)$ that closely approximate $(\alpha, ..., \alpha)$. This step that was comparatively easy in Liouville's proof is not so simple here due to the increased complexity of the auxiliary polynomial.

We illustrate the different approach we will take here for a polynomial with one variable over the rationals.

Example 5.4. Let $P(X) \in \mathbb{Z}[X]$ be a polynomial such that $|P|_\infty \leq B^r$ and let $\text{Ind}(P) = I$ with respect to $(p/q; r)$ for integers p, q, r with $r \geq 1$. We can write

$$P(X) = (qX - p)^{rI} R(X)$$

Where $R(X)$ is a polynomial with integer coefficients by Gauss' lemma. From this it is clear that

$$\max\{|p|_\infty, |q|_\infty\} \leq |P| \leq B^r$$

Which yields

$$\text{Ind}(P) = I \leq \frac{\ln B}{\ln H(p/q)}$$

This result sheds some light onto the hypotheses we'll take for the main result of this section, Roth's lemma. Moreover, $\text{Ind}(P)$ and $\ln H(p, q)$ are inversely proportional, which suggests that we may be able to make $\text{Ind}(P)$ as small as desired for large enough $H(p/q)$. The argument we will make hinges on this.

Before proving Roth's lemma, we will prove a lemma in order to obtain another polynomial with desirable properties. We will need to define the following for the lemma: the differential operator

$$\Delta = \frac{\partial^{i_1 + ... + i_m}}{\partial X_1^{i_1} \cdots X_m^{i_m}}$$

and the order of a differential operator ord(Δ)= $i_1 + \cdots i_m$.

Now, given K a field of characteristic 0, and $\phi_1, ..., \phi_m \in K(X)$, a *Generalized Wronskian determinant* of $\phi_1, ..., \phi_m$ is any determinant of the form

$$det\big((\Delta_i \phi_j)_{1 \leq j, k \leq k}\big)$$

where the operators Δ_i satisfy $\text{ord}(\Delta_i) \leq i - 1$.

Lemma 5.5. Let $\phi_1, ..., \phi_k \in K(X_1, ..., X_m)$ be rational functions over a field of characteristic 0. Then $\phi_1, ..., \phi_k$ are $K-$linearly independent if and only if there exists a nonzero generalized Wronskian of $\phi_1, ..., \phi_k$.

Proof.

We will only prove that if $\phi_1, ..., \phi_k$ are $K-$linearly independent, then there exists a non zero generalized Wronskian for them, as this is all we will need for the proof of Roth's lemma. The proof will be by induction on the number of rational functions, k. For k=1, we have $\det(\Delta_1 \phi) = \phi$ given that Δ_1 must have order 0, so for k=1 the full version of this lemma states that ϕ is $K-$linearly independent iff $\phi \neq 0$.

We now assume that the lemma is true (in the direction we want to prove) for every set of $k-1$ rational functions in $K(X_1, ..., X_m)$, and that $\phi_1, ..., \phi_k$ are $K-$linearly independent. Our aim is to find a non zero generalized Wronskian of these k functions. Now, for a non zero function $\lambda \in K(X_1, ..., X_m)$, any generalized Wronskian of $\lambda\phi_1, ..., \lambda\phi_k$, $\det(\Delta_i(\lambda\phi_j)_{1\leq i,j\leq k})$ is a $K(X_1, ..., X_m)-$linear combination of generalized Wronskians of $\phi_1, ..., \phi_k$, $\det(\Delta_i(\phi_j)_{1\leq i,j\leq k})$. This can be verified using the product rule and the multilinearity of the determinant, and suitably grouping up the same derivatives of $\lambda \in K(X_1, ..., X_m)$ into the same matrix to factor them out. This observation means that proving the existence of a nonzero generalized Wronskian for $\lambda\phi_1, ..., \lambda\phi_k$, will give us the existence of the nonzero Wronskian we are looking for. $\phi_1 \in K(X_1, ..., X_m)$, so we can take $\lambda = 1/\phi_1$. We have thus reduced reduced to the case where $\phi_1 = 1$, and we note that this change preserves the linear independence of the 'new' $\phi_1, ..., \phi_m$.

Consider the $K-$linear span of $\phi_1, ..., \phi_m$

$$V = K\phi_1 + \cdots + K\phi_m \subset K(X_1, ..., X_m)$$

Due to the $K-$linear independence, we have dim V $=k$. Moreover, for every $i \geq 2$, $\phi_i \notin K$ as $\phi_1 = 1$. In particular, $\phi_2 \notin K$ so it's not a constant function, and perhaps after relabelling the variables, we may assume the variable X_1 appears in ϕ_2, implying

$$\frac{\partial\phi_2}{\partial X_1} \neq 0$$

From this observation we define the K-vector subspace of V

$$W = \left\{ \phi \in V \;\Big|\; \frac{\partial\phi}{\partial X_1} = 0 \right\}$$

and define $t := dimW$. $\phi_1 \in W$ and $\phi_2 \notin W$, so $1 \leq t \leq k-1$. W is a vector subspace of V, so we can choose a basis for it: $\{\psi_1, ..., \psi_t\}$, and extend it to a basis $\{\psi_1, ..., \psi_k\}$ of V. From the inequality for t, we notice we can apply the induction hypothesis to the basis of W. This means there are differential operators $\Delta_1^*, ..., \Delta_t^*$ satisfying

$$det(\Delta_i^*\psi_j)_{1\leq i,j\leq t} \neq 0 \qquad \text{and} \qquad ord(\Delta_i^*)_{1\leq i\leq t} \leq i-1$$

Now we need to find similar differential operators for the remaining basis of V. We claim that

$$\frac{\partial\psi_{t+1}}{\partial X_1}, \cdots, \frac{\partial\psi_k}{\partial X_1} \text{ are } K\text{−linearly independent}$$

This assertion follows from the observation that

$$\sum_{i=t+1}^{k} c_i\frac{\partial\psi_i}{\partial X_1} = 0 \iff \sum_{i=t+1}^{k} c_i\psi_i \in W \implies c_{t+1} = \cdots = c_k = 0$$

As they are $K-$linearly independent from the basis of W. Therefore $\{\psi_{t+1}, ..., \psi_k\}$ is a K-basis of V/W, and because $1 \leq dim(W/V) = dimW - dimV \leq k-1$, we can now apply the induction hypothesis obtaining differential operators $\Delta_{t+1}^*, ..., \Delta_k^*$ satisfying

$$det\left(\Delta_i^*\frac{\partial\psi_j}{\partial X_1}\right)_{t+1\leq i,j\leq k} \neq 0 \qquad \text{and} \qquad ord(\Delta_i^*)_{1\leq i\leq t} \leq i-t-1$$

To fit together all the operators we obtained, we now define

$$\Delta_i = \begin{cases} \Delta_i^* & 1 \le i \le t \\ \Delta_i^* \dfrac{\partial}{\partial X_1} & t+1 \le i \le k \end{cases}$$

Note that $\text{ord}(\Delta_i) \le i-1$ for all $1 \le i \le k$. Furthermore, we have

$$\Delta_i \psi_j = \Delta_i \frac{\partial \psi_j}{\partial X_1} = 0 \quad \text{for } 1 \le j \le t \text{ and } t+1 \le i \le k$$

because for $1 \le j \le t$ $\psi_j \in W$. This means that the determinant looks like this

$$det(\Delta_i \psi_j)_{1 \le i,j \le k} = det \left(\begin{array}{c|c} \Delta_i^* \psi_j & \Delta_t^* \psi_0 \\ \hline \mathbf{0} & \Delta_i^* \dfrac{\partial \psi_j}{\partial X_1} \end{array} \right)$$

With the big matrix being separated into 4 major blocks $0 \le i,j \le t$, $t+1 \le i \le k$ and $0 \le j \le k$, $t+1 \le j \le k$ and $0 \le i \le k$ and $t+1 \le i,j \le k$. Now, there is only way to arrange this matrix into 4 submatrices of the size of the aforementioned boxes such that only one out of those 4 matrices has determinant 0, and that is by having those 4 submatrices be precisely the ones shown above. Thus, from Laplace's expansion of the determinant by complementary minors, we obtain

$$det(\Delta_i \psi_j)_{1 \le i,j \le k} = det(\Delta_i^* \psi_j)_{1 \le i,j \le t} \cdot det \left(\Delta_i^* \frac{\partial \psi_j}{\partial X_1} \right)_{t+1 \le i,j \le k} \neq 0 \text{ By induction hypothesis}$$

This doesn't quite complete the proof yet, as this shows there exists a nonzero generalized Wronskian for $\{\psi_1, ..., \psi_k\}$. Now, by construction $\psi_1, ..., \psi_k$ and $\phi_1, ..., \phi_k$ span the same $K-$vector space. Because of this we can write

$$\psi_j = \sum_{l=1}^{k} a_{jl} \phi_l \quad \forall j \in \{1, ..., k\} \quad \text{for some invertible matrix } (a_{jl}) \text{ with coefficients in K}$$

It follows from this that

$$0 \neq det(\Delta_i \psi_j)_{1 \le i,j \le k} = det \left(\sum_l a_{jl} \Delta_i \phi_l \right) = det(a_{jl}) det(\Delta_i \phi_l)_{1 \le i,l \le k}$$

a_{jl} is invertible so $det(\Delta_i \phi_l)_{1 \le i,l \le k} \neq 0$, which concludes the proof.

□

We can now move on to proving Roth's lemma, which implies that our auxiliary polynomial P$(X_1, ..., X_m)$ cannot in fact vanish to high order at $(\beta_1, ..., \beta_m)$. The lemma we just proved is essential as a core idea in the proof of Roth's lemma is to use Wronskians to eliminate a variable and then perform an induction over the number of variables. It is the following lemma that completed the proof of the original version of Roth's theorem, so it is appropriately named after Roth, too.

Proposition 5.6 (Roth's lemma)**.** Let m be a positive integer and let $P \in \overline{\mathbb{Q}}[X_1, ..., X_m]$ be a polynomial with algebraic coefficients and $deg_{X_h}(P) \le r_h$. Let $\beta = (\beta_1, ..., \beta_m)$ be an m-tuple of algebraic numbers. Fix a real number $0 < \eta$ satisfying

$$\frac{r_{h+1}}{r_h} \le \eta^{2^{m-1}} \quad \text{for all } 1 \le h \le m-1 \tag{5.5-i}$$

and

$$\eta^{2^{m-1}} \min_{1\leq h\leq m}\{r_h h(\beta_h)\} \geq h(P) + 2mr_1 \tag{5.5-ii}$$

Then the index of P with respect to $(\beta_1, ..., \beta_m; r_1, ..., r_m)$ satisfies

$$\operatorname{Ind}(P) \leq 2m\eta$$

Remark. It should be noted that we will be able to choose a ver small η. Notice how for $\eta \geq 1/2$ the result of the lemma is trivial as $\operatorname{Ind}(P) \leq m$ (differentiating P in each variable as many times as the highest power of that variable), therefore we may assume $\eta < 1/2$. Now, given η is small, (5.5-i) tells us the degrees r_h are rapidly decreasing, which in turn tells us from (5.5-ii) that the heights $H(\beta_h)$ are very rapidly increasing.

Proof. Let K be a number field such that $d = [K : \mathbb{Q}]$, $P \in K[X_1, ..., X_m]$, and $\beta_i \in K\ \forall i \in \{1, ..., m\}$. The proof is by induction, so we begin by looking at the case for $m = 1$. Let l be the order of vanishing of $P(X)$ at $X = \beta_1$, meaning $P(X) = (X-\beta_1)R(X)$ for some polynomial $R(X)$ with $R(\beta_1) \neq 0$, and $\operatorname{Ind}(P) = l/r_1$ with respect to $(\beta_1; r_1)$. Proposition 3.10, Gelfand's inequality yields

$$\begin{aligned}H(\beta_1)^{r_1 \operatorname{Ind}(P)} =& H(\beta_1)^l = H(X-\beta_1)^l \leq H(X-\beta_1)^l H(Q) \leq H(P)e^r \\ & \text{(Different Heights)}\end{aligned}$$

which implies

$$\operatorname{Ind}(P) \leq \frac{\ln H(P) + r_1}{r_1 \ln H(\beta_1)} \leq \frac{\ln H(P) + 2r_1}{r_1 \ln H(\beta_1)} \leq \eta \quad \text{using 5.5-ii}$$

Which proves the result for $m = 1$.
Note that we obtained a better bound than required for the induction. This will be relevant later to sharpen our estimate when using induction with $m = 1$.

We now assume Roth's lemma is true for polynomials with strictly less than m variables, and will prove it for a polynomial $P(X_1, ..., X_m)$ of m variables with the properties of our auxiliary polynomial. We begin by writing

$$P(X_1, ..., X_m) = \sum_{j=1}^{k} \phi_j(X_1, ..., X_{m-1})\psi_j(X_m) \tag{*}$$

where the ϕ_j and ψ_j are polynomials with coefficients in $\overline{\mathbb{Q}}$. Now, this decomposition of P into the sum of products of lesser degree polynomials may not be unique, so we will choose one that minimizes k. Moreover, note that one possible decomposition is $\psi_j = X_m^j$ for $j \in \{0, 1, ..., r_m\}$, so

$$k \leq r_m + 1 \tag{**}$$

We now show that the polynomials ϕ_j and ψ_j are $\overline{\mathbb{Q}}$–linearly independent which will allow us to apply lemma 5.4. We do this by contradiction, suppose the polynomials ϕ_j are linearly dependent over $\overline{\mathbb{Q}}$, this implies the existence of some constants $c_j \in \overline{\mathbb{Q}}$ such that $\sum_j c_j\phi_j = 0$. Relabelling if necessary, we assume $c_k \neq 0$ which gives the relation

$$\phi_k = -\sum_{j=1}^{k-1} \frac{c_j}{c_k}\phi_j$$

which then gives

$$P(X_1,...,X_m)=\sum_{j=1}^{k}\phi_j\psi_j=\sum_{j=1}^{k-1}\phi_j\psi_j-\sum_{j=1}^{k-1}\frac{c_j}{c_k}\phi_j\psi_k=\sum_{j=1}^{k-1}\phi_j\left(\psi_j-\frac{c_j}{c_k}\psi_k\right)$$

contradicting the minimality of k. We can prove that $\psi_1,...,\psi_k$ are $\overline{\mathbb{Q}}-$linearly independent in exactly the same way, so both $\phi_1,...,\phi_k$ and $\psi_1,...,\psi_k$ are $\overline{\mathbb{Q}}-$linearly independent.

With that groundwork done, we can now define

$$U(X_m):=det\left(\frac{1}{(i-1)!}\frac{\partial^{i-1}}{\partial X_m^{i-1}}\psi_j(X_m)\right)_{1\leq i,j,\leq k}$$

Which is the classical Wronskian determinant of $\psi_1,...,\psi_k$. These polynomials are $\overline{\mathbb{Q}}-$linearly independent, so from lemma 5.4 we know that $U(X_m)\neq 0$. Similarly, we can find differential operators

$$\Delta_i'=\frac{1}{i_1!\cdots i_m!}\frac{\partial^{i_1+\cdots+i_m}}{\partial X_1^{i_1}\cdots\partial X_{m-1}^{i_{m-1}}}$$

with $\text{ord}(\Delta_i')=i_1+\cdots+i_{m-1}\leq i-1\leq k-1\leq r_m$ and a generalized Wronskian determinant satisfying

$$V(X_1,...,X_{m-1}):=det(\Delta_i'\phi_j)_{1\leq i,j\leq k}\neq 0$$

After defining these two poltnomials with coefficients in K, we can now define a third polynomial and exploit that $U(X_m)$ and $V(X_1,...,X_{m-1})$ do not share any variables to compute

$$\begin{aligned}W(X_1,...,X_m)&:=det\left(\Delta_i'\cdot\left(\frac{1}{(j-1)!}\frac{\partial^{j-1}}{\partial X_m^{j-1}}\right)P(X_1,...,X_m)\right)_{1\leq i,j\leq k}\\&=det\left(\Delta_i'\cdot\left(\frac{1}{(j-1)!}\frac{\partial^{j-1}}{\partial X_m^{j-1}}\right)\sum_{h=1}^{k}\phi_h(X_1,...,X_{m-1})\psi_h(X_m)\right)_{1\leq i,j\leq k}\\&=det\left(\sum_{h=1}^{k}\Delta_i'\phi_h\cdot\frac{1}{(j-1)!}\frac{\partial^{j-1}\psi_h}{\partial X_m^{j-1}}\right)_{1\leq i,j\leq k}\\&=det(\Delta_i'\phi_h)_{1\leq i,h\leq k}\cdot det\left(\frac{1}{(j-1)!}\frac{\partial^{j-1}\psi_h}{\partial X_m^{j-1}}\right)_{1\leq h,j\leq k}\\&=V(X_1,...,X_{m-1})U(X_m)\end{aligned}$$

Immediately lemma 5.4 pays off, allowing us to use our auxiliary polynomial P to create a polynomial W with m variables that factorizes into two polynomials with fewer variables than W and P. This will allow us to use Roth's lemma in 1 and $m-1$ variables to bound the indices of U and V, which will lead to a bound for $\text{Ind}(W)$, which will in turn allow us to find a bound for $\text{Ind}(P)$. Before proceeding we note that $W\in K[X_1,...,X_m]$, and that $\partial_{j-1}/\partial X_m^{j-1}$ varies with the columns of the matrix used to define W, which will be relevant later.

Now, for the following section of the proof we will be using the projective height of a polynomial instead of the height we have used previously. This height is defined to be the height of the polynomial's coefficients taken as homogenous coordinates:

$$H_K(P) = \prod_{v \in M_K} \|P\|_v \quad \text{and} \quad h(P) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in M_K} \ln \|P\|_v$$

Where $\|P\|_v$ is defined as usual. Over $\mathbb{Q}$ we have, for example

$$h(6x^2 + 3xy + 12y) = h([6:3:12]) = h([2:1:4]) = ln(4) + \sum_{v \in M_{\mathbb{Q}}^0} ln(1) = ln(4)$$

With this height, and given U and V are polynomials in independent variables, there are no cancellations and we get

$$h(U) + h(V) = h(W)$$

Now we need to estimate the degrees and heights of U and V.

The remaining part of the proof will be a series of smaller results, which together result in the full proof of Roth's lemma.

a) $deg_{X_m}(U) \leq kr_m$ and $deg_{X_j}(V)$ for all $j \in \{1, ..., m-1\}$
Proof. each determinant is of size k as shown in lemma 5.4, and the entries of the determinants for V and U have degree at most r_j with respect to each variable X_j.

b) $h(W) \leq k(h(P) + 2r_1)$
Proof. The determinant is the sum of $k!$ terms, each of which is the product of k polynomials with at most degree r_j for the variable X_j. Therefore, from the definition of W, and using lemma 4.1 we can find a bound for the height of each entry in the determinant of W.

$$H(\Delta_i' \partial_j P(X_1, ..., X_m)) \leq 2^{r_1 + \cdots + r_m} H(P)$$

Where Δ_i' and ∂_j are defined as used in the definition of the polynomial W. Now, It follows from the maximum in the definition of the height that

$$H(W) \leq k! \cdot \left(2^{r_1 + \cdots r_m} H(P)\right)^k$$

as we are adding $k!$ terms, each of them bounded by the bound of an individual entry to the power k (each term in the sum is the product of k polynomials, all bounded by the same term). hence,

$$h(W) \leq k\big(h(P) + (r_1 + \cdots + r_m) \ln 2\big) + \ln k!$$

Now, due to 5.5-i

$$r_1 + \cdots + r_m \leq r_1(1 + \eta' + \cdots + \eta'^{m-1}) \quad \text{with } \eta' = \eta^{2^{m-1}} \text{ for simplicity}$$

we have a finite geometric sum which we can bound by taking $m \geq 2$ and $\eta' \leq (1/2)^2 = 1/4$.

$$r_1 + \cdots + r_m \leq r_1(1 + \eta' + \cdots) \leq \frac{4}{3} r_1$$

On the other hand, $k^k \geq k!$, so

$$\frac{\ln k!}{k} \leq \ln k \leq k-1 \leq r_m \leq \frac{1}{2}r_1$$

This all coalesces into

$$h(W) \leq k\left(h(P) + \left(\frac{4}{3}\ln 2 + \frac{1}{2}\right)r_1\right) \leq k(h(P) + 2r_1)$$

As we wanted. We note however, that we could replace the 2 by $\frac{4}{3}\ln 2 + \frac{1}{2}$, which is smaller, but looks less aesthetic.

We now look to bound the indices of U,V, and W.

c) If Roth's lemma is true for polynomials in $m-1$ or fewer variables, then

$$\operatorname{Ind}(U) \leq k\eta^{2^{m-1}} \quad \text{and} \quad \operatorname{Ind}(V) \leq 2k(m-1)\eta^2$$

with respect to $(\beta_m; r_m)$ and $(\beta_1, ..., \beta_{m-1}; r_1, ..., r_{m-1})$ respectively. Applying lemma 4.5 to this yields

$$\operatorname{Ind}(W) \leq \operatorname{Ind}(U) + \operatorname{Ind}(V) \leq 2k(m-1)\eta^2 + k\eta^{2^{m-1}}$$

with respect to $(\beta_1, ..., \beta_m; r_1, ..., r_m)$. **Proof.** To use our induction hypothesis, we must make sure U and V satisfy hypotheses (5.5-i) and (5.5-ii). We begin with V. V has $m' = m-1$ variables, degrees bounded by $r'_j = kr_j$, and $\eta' = \eta^2$, we now check the conditions.

$$\frac{r'_{j+1}}{r'_j} = \frac{r_{j+1}}{r_j} \leq \eta^{2^{m-1}} = \eta'^{2^{m'-1}} \qquad (5.5\text{-i} \checkmark)$$

The projective height of a polynomial is clearly bigger than 0, so the bound for $h(W)$ from b) works for $h(V)$ too, which gives $h(V) \leq k(h(P) + 2r_1)$. this allows us to write

$$\begin{aligned} r'_j h(\beta_j) = kr_j h(\beta_j) &\geq k\eta'^{-2^{m'-1}}(h(P) + 2mr_1) \qquad (\text{as } \eta^{2^{m-1}} = \eta'^{2^{m'-1}}) \\ &= k\eta'^{-2^{m'-1}}(h(P) + 2r_1) + 2k\eta'^{-2^{m'-1}}(m-1)r_1 \\ &\geq \eta'^{-2^{m'-1}}h(V) + 2\eta'^{-2^{m'-1}}m'r_1 \qquad (\text{k} \geq 1) \end{aligned}$$

For all $j \in \{1, ..., m-1\}$, proving 5.5-ii) for V. A detail that is omitted in this proof is the assertion that the auxiliary polynomial we will be using, P, satisfies 5.5-i) and 5.5-ii). Knowing whether this is true is important, as we just used it to prove V satisfies 5.5-i) and 5.5-ii), and will shortly do the same for U, but it is omitted for now as we will prove it once we move on to proving Roth's theorem using all the results we are proving now. It is better this way as what we have labelled η in this proof will be labelled ω in the proof of the Theorem, and this value will span many of the results we are proving now. We finish up with V by using induction ot show that

$$\begin{aligned} \operatorname{Ind}()(\beta_1, ..., \beta_{m-1}; r_1, ..., r_{m-1})(V) &= k\operatorname{Ind}()(\beta_1, ..., \beta_{m-1}; r'_1, ..., r'_{m-1})(V) \quad \text{By definition} \\ &\leq k(2m'\eta') = 2k(m-1)\eta^2 \end{aligned}$$

We now move on to showing U satisfies 5.5-i) and 5.5-ii).

U has $m'' = 1$ variable, degree bounded by $r'' = kr_m$, and we write $\eta'' = \eta^{2^{m-1}}$. Since U only has 1 variable, 5.5-i) is empty, so we only need to check 5.5-ii)

$$\begin{aligned} h(U) + r'' &\leq k(h(P) + c_1 r_1) + kr_m \leq k(h(P) + 2r_1) \\ &\leq \eta^{2^{m-1}} kr_m h(\beta_m) = \eta'' r'' h(\beta_m) \end{aligned}$$

Where we have used the less aesthetic estimate for $h(W)$ (and $h(U)$) we mentioned before, where $c_1 = \frac{4}{3}\ln 2 + \frac{1}{2} \approx 1.4242$ and thus $c_1 + \eta^{2^{m-1}} \leq c_1 + 1/4 \leq 2$ as $m \geq 2$. Applying now Roth's lemma for a polynomial in 1 variable which we proved to begin the induction we get

$$\operatorname{Ind}((\beta_m; r_m))(U) = k \operatorname{Ind}()(\beta_m; r'')(U) \leq k\eta'' = k\eta^{2^{m-1}}$$

This completes the proof for c), all that is left now is to relate the index of W back to the index of P. By construction, if P vanishes to high order at a point $(\beta_1, ..., \beta_m)$, then the same will be true for every entry in the matrix used to define W, therefore W will also vanish to high order. The following result quantifies this intuition into a concrete statement.

d) With the same notation used previously, we have the following bound for $\operatorname{Ind}(W)$

$$\operatorname{Ind}((W)) \geq \frac{k}{2} \min\left\{\operatorname{Ind}((P)), (\operatorname{Ind}((P)))^2\right\} - k\frac{r_m}{r_{m-1}}$$

Proof. Similarly to b), we begin by estimating the index of the entries of the matrix used to define W, with respect to $(\beta_1, ..., \beta_m; r_1, ..., r_m)$.

$$\begin{aligned} &\operatorname{Ind}\left(\Delta_i'\left(\frac{1}{(j-1)!}\frac{\partial^{j-1}}{\partial X_m^{j-1}}P\right)\right) \\ &\quad = \operatorname{Ind}(\partial_{i_1,...,i_{m-1},j-1}P) \\ &\quad \geq \operatorname{Ind}(P) - \sum_{h=1}^{m-1}\left(\frac{i_h}{r_h}\right) - \frac{j-1}{r_m} \quad \text{(From lemma 4.5)} \\ &\quad \geq \operatorname{Ind}(P) - \frac{i_1 + \cdots + i_{m-1}}{r_{m-1}} - \frac{j-1}{r_m} \quad \text{(From 5.5-i, with } \eta < 1/2) \\ &\quad \geq \operatorname{Ind}(P) - \frac{r_m}{r_{m-1}} - \frac{j-1}{r_m} \quad \text{(as } \operatorname{ord}(\Delta_i') = i_1 + \cdots + i_{m-1} \leq i-1 \leq k-1 \leq r_m) \end{aligned}$$

Now, we know that W is the sum of the product of k polynomials, one from each column, and of the form $\partial_{i_1,...,i_{m-1},j-1}P$ for $j \in \{1, ..., k\}$. Using lemma 4.5 gives a lower bound for the index of W with respect to $(\beta_1, ..., \beta_m; r_1, ..., r_m)$

$$\operatorname{Ind}(W) \geq \min_{\text{k! terms in sum for W}} \{\operatorname{Ind}\big(\text{Product of k polynomials, } \partial_{i_1,...,i_{m-1},j-1}P \quad \forall j\big)\}$$

Applying lemma 4.5 again for the product of k polynomials now yields

$$\operatorname{Ind}(W) \geq \sum_{j=1}^{k} \min_{i_1,...,i_{m-1}} \operatorname{Ind}(\partial_{i_1,...,i_{m-1},j-1}P)$$

Substituting in the lower bound for $\partial_{i_1,\ldots,i_{m-1},j-1}P$ we obtained previously, only when it is positive (The index is nonnegative) results in the following inequality

$$\begin{aligned}\operatorname{Ind}(W) &\geq \sum_{j=1}^{k} \max\left\{\operatorname{Ind}(P) - \frac{r_m}{r_{m-1}} - \frac{j-1}{r_m}, 0\right\} \\ &\geq \sum_{j=1}^{k} \max\left\{\operatorname{Ind}(P) - \frac{j-1}{r_m}, 0\right\} - \frac{kr_m}{r_{m-1}}\end{aligned}$$

All that is left now is finding a lower bound for

$$\sum_{j=1}^{k} \max\left\{\operatorname{Ind}(P) - \frac{j-1}{r_m}, 0\right\}$$

As a function of $\operatorname{Ind}(P)$, we will consider two cases.

Case 1. $\operatorname{Ind}(P) \geq \dfrac{k-1}{r_m}$

This case is simple, as it suffices to compute

$$\sum_{j=1}^{k} \left(\operatorname{Ind}(P) - \frac{j-1}{r_m}\right) = k \operatorname{Ind}(P) - \frac{(k-1)k}{2r_m} \geq \frac{k}{2}\operatorname{Ind}(P)$$

Case 2. $\operatorname{Ind}(P) \leq \dfrac{k-1}{r_m}$

We begin by defining $N = \lfloor r_m \operatorname{Ind}(P) \rfloor$. From our case we have $N \leq k-1$, and if we assume that j_0 is the smallest j such that $\operatorname{Ind}(P) - \dfrac{j_0 - 1}{r_m} < 0$, we get $N \leq r_m \operatorname{Ind}(P) < j_0 - 1 \implies N + 1 = j_0 - 1$, meaning for this case we only need to find a bound for

$$\sum_{j=1}^{k} \max\left\{\operatorname{Ind}(P) - \frac{j-1}{r_m}, 0\right\} = \sum_{j=1}^{N+1} \left(\operatorname{Ind}(P) - \frac{j-1}{r_m}\right)$$

This sum becomes

$$\begin{aligned}&\sum_{j=1}^{N+1} \left(\operatorname{Ind}(P) - \frac{j-1}{r_m}\right) \\ &\quad= (N+1)\operatorname{Ind}(P) - \frac{N(N+1)}{2r_m} \\ &\quad= (N+1)\left(\operatorname{Ind}(P) - \frac{\lfloor r_m \operatorname{Ind}(P) \rfloor}{2r_m}\right) \quad \text{From the definition of N} \\ &\quad\geq (N+1)\frac{1}{2}\operatorname{Ind}(P) \geq r_m \operatorname{Ind}(P)\frac{\operatorname{Ind}(P)}{2} \quad \text{Definition of N} \\ &\quad\geq \frac{k}{2}\operatorname{Ind}(P)^2 \quad \text{assuming } k \leq r_m\end{aligned}$$

We know that $k \leq r_m + 1$ though, so we now check the case assuming $k = r_m + 1$. Substituting $k - 1 = r_m$ we obtain

$$q(N) := \sum_{j=1}^{N+1} \left(\operatorname{Ind}(P) - \frac{j-1}{r_m} \right) = (N+1) \operatorname{Ind}(P) - \frac{N(N+1)}{2(k-1)}$$

$q(N)$ is a quadratic function of N with negative leading coefficient, meaning it will achieve it's minimum value for either the minimum or maximum value of N. By definition, we have the following bounds for N

$$(k-1) \operatorname{Ind}(P) - 1 \leq N = \lfloor (k-1) \operatorname{Ind}(P) \rfloor \leq (k-1) \operatorname{Ind}(P)$$

We now use these bounds to compute the minimum and find a lower bound for $q(N)$.

$$q((k-1) \operatorname{Ind}(P) - 1) = q((k-1) \operatorname{Ind}(P)) = \frac{(k-1) \operatorname{Ind}(P)^2 + \operatorname{Ind}(P)}{2}$$

Since $q(N)$ takes on the same value at both extremes we know it's the minimum, now, since $\operatorname{Ind}(P) \leq 1$ we have

$$q(N) \geq \frac{(k-1) \operatorname{Ind}(P)^2 + \operatorname{Ind}(P)}{2} \geq \frac{k \operatorname{Ind}(P)^2}{2}$$

This completes the second case.

Consequently, and after rearranging, we find the following bound for $\operatorname{Ind}(W)$:

$$\operatorname{Ind}(W) + \frac{k r_m}{r_{m-1}} \geq \sum_{j=1}^{k} \max \left\{ \operatorname{Ind}(P) - \frac{j-1}{r_m}, 0 \right\} \geq \frac{k}{2} \min\{ \operatorname{Ind}(P), \operatorname{Ind}(P)^2 \}$$

As already mentioned, $\operatorname{Ind}(P) \leq m$, therefore

$$\operatorname{Ind}(W) + \frac{k r_m}{r_{m-1}} \geq \frac{k}{2} \min\{ \operatorname{Ind}(P), \operatorname{Ind}(P)^2 \} \geq \frac{k \operatorname{Ind}(P)^2}{2m}.$$

From c) we also have an upper bound for $\operatorname{Ind}(W)$, which implies that

$$\begin{aligned} \operatorname{Ind}(W) + \frac{k r_m}{r_{m-1}} &\leq 2k(m-1)\eta^2 + k\eta^{2^{m-1}} + \frac{k r_m}{r_{m-1}} \\ &\leq k\left(2(m-1)\eta^2 + 2\eta^{2^{m-1}}\right) \quad \text{hypothesis 5.5-i)!} \\ &\leq k(2m\eta^2) \end{aligned}$$

Together, both bounds yield

$$\frac{k \operatorname{Ind}(P)^2}{2m} \leq k(2m\eta^2) \implies \operatorname{Ind}(P)^2 \leq (2m\eta)^2$$

The index is nonnegative, so we finally obtain

$$\operatorname{Ind}(P) \leq 2m\eta$$

□

5.4 Completion of the Proof of Roth's Theorem

We have finally assembled all the necessary results needed to complete the proof of Roth's theorem, although, more precisely, we will prove Theorem 4.9, which was proven to be equivalent to Roth's Theorem. For convenience, and aesthetic purposes, we will restate the result in this section, however, instead of ε as before we will use δ.

Theorem 5.7. Let K be a number field, let $S \subset M_K$ be a finite set of absolute values on K with each absolute value extended in some way to $\overline{K}$. Given $\alpha \in \overline{K}$ and $\delta > 0$, suppose that

$$\zeta : S \to [0,1] \quad \text{is a function satisfying} \quad \sum_{v \in S} \zeta_v = 1$$

Then there are only finitely many $\beta \in K$ such that

$$\|\beta - \alpha\|_v \leq \frac{1}{H_K(\beta)^{(2+\delta)\zeta_v}} \quad \forall v \in S \tag{2}$$

Proof. We will assume there are infinitely many solutions to (2) and derive a contradiction. The basic scheme of the proof is to pic a large integer m, and suppose there are solutions $\beta_1, ..., \beta_m$ to (2) satisfying certain conditions. We then use the results we have proven in this chapter to produce a polynomial with certain properties, and show it's index is both bigger and smaller than $m\varepsilon$ with respect to $(\beta_1, ..., \beta_m; r_1, ..., r_m)$, providing us with our desired contradiction.

Due to the numerous conditions and hypotheses employed in the results of this section which will be used in this proof, we will briefly list them here again for convenience. Each constant is defined in the result it belongs to.

(5.1 hypothesis)

(a) Given $\varepsilon > 0$, choose m such that $e^{m\varepsilon^2/4} > 2[\mathbb{Q}(\alpha) : \mathbb{Q}] = 2d$

(5.1 results)

(a) $\deg_{X_h}(P) \leq r_h \ \ \forall h \in \{1, ..., m\}$

(b) $\text{Ind}(P) \geq \frac{m}{2}(1 - 2\varepsilon)$ w.r.t $(\alpha, ..., \alpha; r_1, ..., r_m)$

(c) $|P|_\infty \leq B(\alpha)^{r_1 + \cdots + r_m}$

(5.2 hypotheses)

(a) $0 < \varepsilon < \delta/28$

(b) $\|\beta_h - \alpha\|_v \leq \dfrac{1}{H_K(\beta_h)^{(2+\delta)\zeta_v}}$

(c) $D := \min_{1 \leq h \leq m} \{\text{H}(\beta_h)^{r_h}\} \leq \max_{1 \leq h \leq m} \{\text{H}(\beta_h)^{r_h}\} \leq D^{1+\varepsilon}$

(d) $C(\alpha, \delta) \leq \min_{1 \leq h \leq m} \{H(\beta_h)\}$

(5.2 result)

(a) $\text{Ind}(P) \geq \varepsilon m$ w.r.t $(\beta_1, ..., \beta_m; r_1, ..., r_m)$

(5.5 hypotheses)

(a) $r_{h+1} \leq \omega r_h \ \ \forall h \in \{1, ..., m-1\}$

(b) $\ln |P|_\infty + 2mr_1 \leq \omega \ln D$

1. (5.5 result)

   (a) $\text{Ind}(P) \leq 2m\omega^{2^{-m+1}}$ w.r.t $(\beta_1, ..., \beta_m; r_1, ..., r_m)$

We assume now that there are infinitely many solutions to (2). Decreasing δ only serves to make the Theorem stronger, so we can assume that $0 < \delta < 1$. We now choose the quantities in the list in the following order:

1. Choose ε such that $0 < \varepsilon < \delta/28$, which implies $0 < \varepsilon < 1/28$. Therefore ε satisfies the hypotheses of 5.1a) and 5.2a.

2. Choose a positive integer m such that $e^{m\varepsilon^2/4} > 2[\mathbb{Q}(\alpha) : \mathbb{Q}] = 2d$. Then all the hypotheses of 5.1 are true, and we can define $\omega := \omega(m, \varepsilon) = (\varepsilon/4)^{2^{m-1}}$, which implies $2\omega^{2^{-m+1}} = \varepsilon/2 < \varepsilon$

3. (2) has infinitely many solutions in K by assumption, but since K only has finitely many elements of bounded height by Theorem 3.4, so we can find a solution β_1 with a height as large as we desire. In particular:

$$H(\beta_1) \geq C(\alpha, \delta) \quad \ln H(\beta_1) \geq \frac{m(\ln B(\alpha) + 2)}{\omega}$$

4. We then use β_1 to choose $m-1$ more solutions to (2) successively, namely $\beta_2, ..., \beta_m$ that satisfy

$$H(\beta_{h+1})^\omega \geq H(\beta_h)^2 \ \ \forall h \in \{1, ..., m-1\}$$

   Since $\omega < 1$, we have $H(\beta_{h+1}) \geq H(\beta_h) \ \forall h \in \{1, ..., m-1\}$, which satisfies hypothesis 5.2d). We note that by choosing β_h's satisfying (2), have also verified that hypothesis 5.2b) is true.

5. Choose a positive integer r_1 such that $H(\beta_1)^{\omega r_1} \geq H(\beta_m)^2$

6. Our aim is to choose positive integers $r_2, ..., r_m$ so that all the $H(\beta_h)^{r_h}$'s are approximately equal. Using r_1 as our 'base point', we can define

$$r_h = \left\lceil \frac{r_1 \ln H(\beta_1)}{\ln H(\beta_h)} \right\rceil$$

   Where $\lceil x \rceil$ is the ceiling function. We want to verify these choices of r_h's satisfy hypothesis 5.2c)

$$\begin{aligned} r_1 \ln H(\beta_1) &\leq r_h \ln H(\beta_h) \ \ \text{By definition of } r_h \text{ and } \lceil x \rceil \\ &\leq r_1 \ln H(\beta_1) + \ln H(\beta_h) \ \ \text{As } r_h \leq 1 + r_1 \ln H(\beta_1)/\ln H(\beta_h) \\ &\leq r_1 \ln H(\beta_1) + \ln H(\beta_m) \ \ \text{As } H(\beta_h) \text{ is increasing over h} \\ &\leq r_1 \ln H(\beta_1) + \omega r_1 \ln H(\beta_1)/2 \ \ \text{From our choice of } r_1 \\ &\leq (1+\varepsilon) r_1 \ln H(\beta_1) \ \ \text{From our definition of } \omega \text{ and } \varepsilon << 4 \end{aligned}$$

Taking exponents gives

$$H(\beta_1)^{r_1} \leq H(\beta_h)^{r_h} \leq H(\beta_1)^{r_1(1+\varepsilon)} \ \ \forall h \in \{1, ..., m\}$$

Therefore $D = H(\beta_1)^{r_1}$ and $\max\limits_{1 \leq h \leq m} \{H(\beta_h)^{r_h}\} \leq D^{(1+\varepsilon)}$, satisfying hypothesis 5.2c). Now, from the definition of r_h we can compute

$$\begin{aligned} \frac{r_{h+1}}{r_h} &= \frac{\left\lceil \dfrac{r_1 \ln H(\beta_1)}{\ln H(\beta_{h+1})} \right\rceil}{\left\lceil \dfrac{r_1 \ln H(\beta_1)}{\ln H(\beta_h)} \right\rceil} \\ &\leq \left(\frac{r_1 \ln H(\beta_1)}{\ln H(\beta_{h+1})} + 1 \right) \Big/ \left(\frac{r_1 \ln H(\beta_1)}{\ln H(\beta_h)} \right) \\ &= \frac{\ln H(\beta_1)}{\ln H(\beta_{h+1})} + \frac{r_1 \ln H(\beta_h)}{\ln H(\beta_1)} \\ &\leq \frac{\omega}{2} + \frac{\omega}{2} = \omega \ \text{ From our choice of } r_1 \text{ and } \beta_h\text{'s in (4) and (5)} \end{aligned}$$

Hence $r_{h+1} \leq \omega r_h$, which verifies hypothesis 5.5a).
We have now verified every hypothesis except for 5.5b), so we can use propositions 5.1 and 5.2.

7. Proposition 5.1 provides an auxiliary Polynomial $P(X_1, ..., X_m) \in \mathbb{Z}[X_1, ..., X_m]$ satisfying the three results 5.1a), 5.1b) and 5.1c).

8. Now that we have the polynomial, applying proposition 5.2 yields

$$\operatorname{Ind}(P) \geq m\varepsilon \quad \text{with respect to } (\beta_1, ..., \beta_m; r_1, ..., r_m)$$

9. All that's left is verifying hypothesis 5.5b) to apply Roth's lemma to the auxiliary polynomial as well. We identify $\omega = \eta^{2^{m-1}}$ and use the fact that

$$\ln D = r_1 \ln H(\beta_1) \ \text{ and } \ r_1 = \max_h \{r_h\} \ \text{ As } H(\beta_h) \text{ increases in h}$$

to compute

$$\begin{aligned} \frac{\ln |P|_\infty + 2mr_1}{\ln D} &\leq \frac{(r_1 + \cdots + r_m) \ln B(\alpha) + 2mr_1}{\ln D} \ \text{ From result 5.1c)} \\ &\leq \frac{r_1(\ln |P|_\infty + 2)}{\ln H()\beta_1} \\ &\leq \omega \ \text{ From our choice of } \beta_1 \text{ back in (3)} \end{aligned}$$

This completes the verification of all the hypotheses and allows us to apply Roth's lemma to the auxiliary polynomial with $\eta = \omega^{2^{-m+1}} = \varepsilon/4$. This yields

$$\operatorname{Ind}(P) \leq 2m\eta = m\varepsilon/2 \quad \text{with respect to } (\beta_1, ..., \beta_m; r_1, ..., r_m)$$

However, (8) and (9) together give an impossible bound for $\operatorname{Ind}(P)$ with respect to $(\beta_1, ..., \beta_m; r_1, ..., r_m)$

$$m\varepsilon \leq \operatorname{Ind}(P) \leq m\varepsilon/2$$

Contradicting the hypothesis that there are infinitely many $\beta \in K$ that satisfy (2). This proves Theorem 4.9, which, by Theorem 4.10, is equivalent to a proof of Roth's Theorem.

5.5 On the effectivity of the proof

Having now proved Roth's Theorem, we delve into it's significance to Diophantine Approximation. Roth's Thorem is purely *qualitative*, it simply tells us there are finitely many solutions, but doesn't provide any details regarding those solutions, if there are any. This is due to the assumptions that engineer the proof: we begin by assuming there is a solution β_1 whose height is as large as needed, then β_2 has a much larger height that depends on the height of the previous solution, and so on. We do not have a way to know anything about β_1, much less the solutions that come after it that depend on it, and due to assuming there is a solution, there may not be any, and we would still have a finite amount of solutions.

It is in fact still an open problem to make Roth's proof *effective*, that is, give an effective algorithm to compute all the elements of the set of solutions. On the other hand, ever since Roth, progress has been made on finding *quantitative* results, which means giving an upper bound for the number of solutions.

6 Applications of Roth's theorem: the unit equation

Here we begin to reap the rewards for proving this version of Roth's theorem, but first of all we must define the ring of S-integers.

Definition 6.1 (Ring of S-integers)**.** Given $S \subset M_K$ a set of absolute values containing the archimedian absolute values M_K^∞, then the ring of S-integers of K is defined as

$$R_S = \{x \in K \mid \ |x|_v \leq 1 \ \ \forall v \in M_K \smallsetminus S\}$$

We note that this definition comes from generalizing the ring of integers of K, which comes from the observation that the non-archimedean absolute values M_K^0 are those corresponding to the prime ideals. The ring of integers of K can be characterized with absolute values as

$$R_K = \{x \in K \mid \ |x|_v \leq 1 \ \ \forall v \in M_K^0\}$$

With this we are ready to show that the two variable S-unit equation has finitely many solutions.

Theorem 6.2 (Siegel, Mahler)**.** Let $K/\mathbb{Q}$ be a field extension, let $S \subset M_K$ be a finite set of absolute values on K that includes all the archimedean absolute values, and let R_S be the ring of S-integers of K. Then the S-unit equation

$$U + V = 1$$

Has only finitely many solutions in S-units $U, V \in R_S^*$
The proof will rely on Roth's theorem, so we need a way to find a link between the solutions of the unit equation and Diophantine approximation. We begin by noticing that by definition, $U, V \in R_S^*$ means there is some absolute value $w \in S$ such that both $|U|_w$ and $|V|_w$ are large (≥ 1). $U, V \neq 0$, so we can rearrange the equation into

$$\left|\frac{U}{V} - 1\right|_w \leq \frac{1}{|V|_w}$$

Meaning U/V is a good approximation of 1. This of course doesn't contradict Roth's theorem, we must use properties of R_S^* and manipulate this equality some more. By [2, Theorem C.3.3] R_S^* is finitely generated, so we can substitute $U = aX^m$ and $V = bY^m$ for some integer m. We will then show that X/Y is almost as close to $\sqrt[m]{\frac{-b}{a}}$ as U/V is to 1, while $H_K(X/Y)$ is approximately $H_K(U/V)^{1/m}$. This means that we can make the height smaller without significantly changing the approximation distance, this allows us to take m as large as necessary to contradict Roth's theorem.
Proof.(of Theorem 6.1) Suppose there are infinitely many solutions $U, V \in R_S^*$ to the S-unit equation $U + V = 1$. Let $s = \#S$ be the number of absolute values in S, and fix a positive integer $m = 2s + 1$. R_S^* is finitely generated, so it's quotient group $R_S^*/R_S^{*^m}$ is also finite and we can fix a set of coset representatives, A. Hence, every element of R_S^* can be written as an element of A multiplied by a member of $R_S^{*^m}$, that is, an m-th power. This defines a map

$$\begin{aligned} \{(U, V) \in R_S^* \times R_S^* \mid \ U + V = 1\} &\to A \times A \\ (U, V) &\longmapsto (a, b) \end{aligned}$$

With $U/a, V/b \in R_S^{*^m}$. However the set on the left is the one we assume is infinite, but clearly $A \times A$ is finite, so by the pidgeonhole principle there must be some $(a, b) \in A \times A$ that corresponds to infinitely many (U, V)'s. Writing $U/a = X^m$ and $V/b = Y^m$, which is equivalent to saying there exist $a, b \in A$ such that

$$aX^m + bY^m = 1$$

has infinitely many solutions $X, Y \in R_S^*$. Now, the set S has finitely many elements, and we can assign each solution to an absolute value by assigning (X, Y) to the absolute value w that maximizes $\|Y\|_w$. Since we have infinite solutions and s is finite, there's at least an absolute value corresponding to an infinite number of solutions, we call it w, and fix $\alpha = \sqrt[m]{\frac{-b}{a}}$. Then

$$\frac{1}{aY^m} = \frac{aX^m + bY^m}{aY^m} = \frac{X^m}{Y^m} + \frac{b}{a} = \frac{X^m}{Y^m} - \alpha^m = \prod_{\zeta \in \mu_m} \left(\frac{X}{Y} - \zeta\alpha \right)$$

Where μ_m is the set of all the m^{th} roots of unity. By taking $\|Y\|$ to be large, we are ensuring that at least one of the factors in the product is small. Moreover, if we consider geometrically where the points $\zeta\alpha$ lie on the complex plane, it makes sense to claim only one (at the absolute most two) of the factors $|X/Y - \zeta\alpha|$ is 'small'. To see this, let $\zeta, \zeta' \in \mu_m$ be distinct m^{th} roots of unity. Then, using the triangle inequality we get

$$\left| \frac{X}{Y} - \zeta\alpha \right|_w + \left| \frac{X}{Y} - \zeta'\alpha \right|_w \geq |\zeta\alpha - \zeta'\alpha|_w \geq C_1$$

Where we define this constant $C_1 = C_1(K, S, m)$ independently of X and Y, as they only depend on the smallest 'distance' between tw of the $\zeta\alpha$. Furthermore, despite depending on α, since α comes from our choice of S and m, C_1 is well defined. It follows then that

$$\frac{1}{|aY^m|_w} = \prod_{\zeta \in \mu_m} \left| \frac{X}{Y} - \zeta\alpha \right|_w \geq \left(\min_{\zeta \in \mu_m} \left| \frac{X}{Y} - \zeta\alpha \right|_w \right) \cdot \left(\frac{C_1}{2} \right)^{m-1}$$

Since at most two terms of the sum can be equal to $C_1/2$, which only happens when it equals $\min_{\zeta \in \mu_m} \left| \frac{X}{Y} - \zeta\alpha \right|_w$, therefore

$$\frac{1}{\|Y\|_w^m} \geq C_2 \min_{\zeta \in \mu_m} \left\| \left(\frac{X}{Y} - \zeta\alpha \right) \right\|_w$$

We now use the pidgeon hole principle in a similar way as before. The pidgeons are the solutions (X, Y), and we place them into th epidgeonholes that are the m roots of unity, according to which root of unity $\zeta \in \mu_m$ minimizes $\|(X/Y) - \zeta\alpha\|_w$. We call the m^{th} degree unity corresponding to the pidgeonhole with infinite solutions ξ. Therefore

$$\frac{1}{\|Y\|_w^m} \geq C_2 \left\| \left(\frac{X}{Y} - \xi\alpha \right) \right\|_w$$

Which shows that X/Y is a good approximation to $\xi\alpha$, all that is left now is finding a way to relate $\|Y\|_w$ to the height of X/Y.

We chose w in order to maximize $\|Y\|_v$, and given that $\|Y\|_v = 1 \;\; \forall v \notin S$ we have

$$\|Y\|_v = \max_{v \in S} \|Y\|_v \geq \Big(\prod_{v \in S} \|Y\|_v \Big)^{1/s} = \Big(\prod_{v \in M_K} \max\{1, \|Y\|_v\} \Big)^{1/s} = H_K(Y)^{1/s}$$

Using now that

$$H(x+y) \leq 2H(x)H(y) \qquad H(xy) \leq H(x)H(y)$$

and that (X,Y) is a solution of $aX^m + bY^m = 1$, we can compute

$$H_K\left(\frac{X^m}{Y^m}\right) = H_K\left(\frac{1}{aY^m} - \frac{b}{a}\right) \leq 2^{[K:\mathbb{Q}]} H_K\left(\frac{1}{aY^m}\right) H_K\left(\frac{b}{a}\right)$$

$$\leq 2^{[K:\mathbb{Q}]} H_K\left(\frac{1}{Y^m}\right) H_K\left(\frac{b}{a}\right) H_K\left(\frac{1}{a}\right)$$

Using that $H_K(T^m) = H_K(T)^m$ and taking m^{th} roots we obtain another constant $C_3(K,S,m)$ such that

$$H_K(X/Y) \leq C_3 H_K(1/Y) = C_3 H_K(Y)$$

Together with the other bound we get

$$\|Y\|_w \geq C_3^{-1/s} H_K(X/Y)^{1/s}$$

And combining this now with the inequality showing X/Y is a good approximation to $\xi\alpha$ we get

$$\frac{1/(C_2 C_4^m)}{H_K(X/Y)^{m/s}} \geq \left\| \left(\frac{X}{Y} - \xi\alpha\right) \right\|_w$$

By assumption, we have infinitely many $X, Y \in R_S^*$ satisfying this inequality, but since we chose $m = 2s+1$, $m/s = 2 + 1/s > 2$, so Roth's theorem tells us there are only finitely many solutions to this inequality in K. This contradicts our assumption, hence, there's only finitely many solutions to the unit equation over R_S^*.

To complement this theorem, we will announce a much stronger result that provides a quantitative outlook on the unit equation, thanks to Evertse.

Theorem 6.3. Let $K/\mathbb{Q}$ be a field extension, let $S \subset M_K$ be a finite set of absolute values on K that includes all archimedean absolute values, and let R_S be the ring of S-integers of K. Then for any $A, B \in K^*$ the S-unit equation

$$AU + BV = 1$$

has at most $3 \cdot 7^{[K:\mathbb{Q}]+2\#S}$ solutions (U,V) over R_S^*.

References

[1] Enrico Bombieri, Walter Gubler; *Heights in Diophantine Geometry*; Cambridge University Press

[2] Marc Hindry, Joseph H.Silverman; *Diophantine Geometry*,An introduction; Springer

[3] S. Lang; *Algebra*,revised 3rd edition; Springer Revised third edition.

[4] G.H Hardy, E.M Wright; *An Introduction to the Theory of Numbers*, 6th edition; Oxford Mathematics

[5] JÃ¼rgen Neukirch; *Algebraic Number Theory*; Springer

[6] B.Edixhoven, J.-H. Evertse; *Diophantine Approximation and Abelian Varieties* ,Introductory lectures; Springer-Verlag

[7] Umberto Zannier; *Lecture notes on Diophantine Analysis*; Edizioni della Normale

[8] S.Lang; *Fundamentals of Diophantine Geometry*; Springer Science+Business Media, LLC

[9] K.F.Roth; *Rational Approximations to Algebraic Numbers*; Mathematika, a journal of pure and applied mathematics