An Introduction to Diophantine Geometry

Ziyang Gao
Leibniz Universität Hannover

2022-03-06

Contents

Chapter 1

Heights on Projective and Affine Spaces

1.1 Absolute values

1.1.1 Basic notions

Definition 1.1.1. *An **absolute value** on a field K is a real valued function $|\cdot|$ on K such that*

(a) $|x| \geq 0$ *and* $|x| = 0$ *if and only if* $x = 0$.

(b) $|xy| = |x||y|$.

(c) $|x + y| \leq |x| + |y|$ *(triangle inequality).*

If furthermore $|\cdot|$ satisfies instead of (c) the stronger condition

(c') $|x + y| \leq \max\{|x|, |y|\}$ *(ultrametric triangle inequality),*

*then it is called **non-archimedean**. If (c') fails to hold for some $x, y \in K$, then the absolute value is called **archimedean**.*

Example 1.1.2. *(i) The **trivial absolute value**: $|0| = 0$ and $|x| = 1$ for all $x \in K^*$.*

(ii) $K = \mathbb{Q}$

- *An archimedean absolute value defined by*

$$|x|_\infty = \max\{x, -x\}.$$

- *A non-archimedean absolute value for each prime number p defined as follows. For any nonzero rational number $x \in \mathbb{Q}$, there exists a unique integer $\mathrm{ord}_p(x)$ such that x can be written in the form*

$$x = p^{\mathrm{ord}_p(x)} \frac{a}{b} \quad \text{with } a, b \in \mathbb{Z} \text{ and } p \nmid ab.$$

*If $x = 0$, then we set $\mathrm{ord}_p(x) = +\infty$. The **$p$-adic absolute value** of $x \in \mathbb{Q}$ is the quantity*

$$|x|_p = p^{-\mathrm{ord}_p(x)}.$$

Intuitively, x is p-adically small if it is divisible by a large power of p.

Each absolute value $|\cdot|$ on K induces a topology via the metric defined by $\mathrm{disc}(x,y) = |x-y|$. If two absolute values define the same topology, they are called *equivalent*. Here is a basic property.

Proposition 1.1.3. *Two absolute values $|\cdot|_1$ and $|\cdot|_2$ are equivalent if and only if there exists a positive real number s such that*

$$|x|_1 = |x|_2^s$$

for each $x \in K$.

In practice, it is more convenient to study equivalence classes of absolute values.

Definition 1.1.4. *A **place** v is an equivalent class of non-trivial absolute values. By $|\cdot|_v$ we denote an absolute value in the equivalence class determined by the place v.*

*We say that a place v is **(non-)archimedean** if $|\cdot|_v$ is.*

As an example, $\mathbb{Q}$ has a unique archimedean place and there is a natural bijection

$$\{\text{non-archimedean places of } \mathbb{Q}\} \leftrightarrow \{\text{all prime numbers}\};$$

see Example 1.1.2.(ii) for $|\cdot|_v$ with each place v of $\mathbb{Q}$.

Consider a field extension K/K_0. For a place v of K, the restriction of $|\cdot|_v$ to K_0 is an absolute value of K_0, and hence is a representative of a place of K_0. We write $v|v_0$ if and only if the restriction of $|\cdot|_v$ to K_0 is a representative of $v_0 \in M_{K_0}$. In this case, we say that v *divides* v_0 or v *lies over* v_0 or v *extends* v_0.

Before moving on, let us look at the example of an arbitrary number field K.

Example 1.1.5. *By definition, $K/\mathbb{Q}$ is a finite field extension, and hence any place v of K lies over some place of $\mathbb{Q}$. There are two possibilities: either $v|p$ for a prime number p, or $v|\infty$ for the unique archimedean place ∞ of $\mathbb{Q}$. It can be then checked that*

$$\{\textit{non-archimedean places of } K\} \leftrightarrow \{\textit{all prime ideals of } \mathcal{O}_K\}$$

and

$$\{\textit{archimedean places of } K\} \leftrightarrow \{\textit{equivalence classes of embeddings } K \hookrightarrow \mathbb{C}\}.^{[1]}$$

We will come back to this with a more precise description of the bijections in Example 1.1.11.

We close this subsection with the following discussion. Let K be a field with a non-archimedean place v. The *valuation ring* of v is defined to be

$$R_v := \{x \in K : |x|_v \leq 1\}.$$

The definition is clearly independent of the choice of $|\cdot|_v$. It can be checked that R_v is local ring with unique maximal ideal $\mathfrak{m}_v := \{x \in K : |x|_v < 1\}$. The *residue field* $k(v)$ is defined to be $R_v/\mathfrak{m}_v$. The quotient map $R_v \to k(v),\ x \mapsto \overline{x}$ is called the *reduction*.

For example when $K = \mathbb{Q}$ and v corresponds to the prime number p, we have $R_v = \{x \in \mathbb{Q} : p^{-\mathrm{ord}_p(x)} \leq 1\} = \{x \in \mathbb{Q} : \mathrm{ord}_p(x) \geq 0\} = \{\frac{a}{b} : a \text{ and } b \text{ coprime},\ p \nmid b\}$ and $\mathfrak{m}_v = \{x \in \mathbb{Q} : \mathrm{ord}_p(x) > 0\} = \{\frac{a}{b} : a \text{ and } b \text{ coprime},\ p|a\} = pR_v$. The residue field is $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

The place v is called *discrete* if the value group $|K^*|_v$ is cyclic. Then $\mathfrak{m}_v$ is a principal ideal and any principal generator is called a *local parameter*. This is the case for the example above[2] and a local parameter is p.

[1] Two embeddings $\sigma_1, \sigma_2 \colon K \hookrightarrow \mathbb{C}$ are equivalent if and only if they are conjugate (*i.e.* $\sigma_2(x) = \overline{\sigma_1(x)}$ for all $x \in K$).

[2] This holds true for any number field and a non-archimedean place.

1.1.2 Normalized absolute values

For each place v of K, we would like assign a well-chosen absolute value. In this subsection we do this.

Definition-Proposition 1.1.6. *For a place v of K, there exists a unique (up to isometric isomorphisms) pair (K_v, w) with K_v/K an extension and w a place of K_v satisfying the following properties:*

(a) $w|v$.

(b) The topology of K_v induced by w is complete.

(c) K is dense in K_v in the above topology.

*This K_v is called the **completion** of K with respect to v. By abuse of notation, we shall denote the unique place w also by v.*

As an example, the field $\mathbb{Q}_p$ of p-adic numbers is the completion of $\mathbb{Q}$ with respect to the place p, and the completion of $\mathbb{Q}$ with respect to the archimedean place is $\mathbb{R}$. In general, we have:

Theorem 1.1.7 (Ostrowski)**.** *The only complete archimedean fields are $\mathbb{R}$ and $\mathbb{C}$.*

An elementary result of the local and global degrees is the following equality. It can be proved using the primitive element theorem.

Lemma 1.1.8. *Let K_0 be a field with a place v_0, and let K/K_0 be a finite separable extension. Then*

$$\sum_{v|v_0}[K_v : K_{0,v_0}] = [K : K_0].$$

With these preparations in hand, we are ready to state the following result about the uniqueness of the extension of absolute values.

Proposition 1.1.9. *Let K_0 be a field which is complete with respect to an absolute value $|\cdot|_{v_0}$ (i.e. $K_0 = K_{0,v_0}$) and let K/K_0 be a finite extension. Then there exists a unique extension of $|\cdot|_{v_0}$ to an absolute value $|\cdot|_v$ of K. Furthermore, for each $x \in K$, we have*

$$|x|_v = |N_{K/K_0}(x)|_{v_0}^{1/[K:K_0]}$$

where N_{K/K_0} is the norm. Moreover, K is complete with respect to $|\cdot|_v$, i.e. $K = K_v$.

Inspired by this proposition, we make the following constructions. Let K_0 be a field with a non-trivial absolute value $|\cdot|_{v_0}$. Let K/K_0 be a finite separate extension with a place v such that $v|v_0$. For any $x \in K$, define

$$|x|_v := |N_{K_v/K_{0,v_0}}(x)|_{v_0}^{1/[K_v:K_{0,v_0}]} \tag{1.1.1}$$

and

$$\|x\|_v := |N_{K_v/K_{0,v_0}}(x)|_{v_0}. \tag{1.1.2}$$

The following statements are easy to verify.

- The $|\cdot|_v$ defined above is an absolute value representing v by Proposition 1.1.9.

- The $\|\cdot\|_v$ defined above is an absolute value representing v *unless* $K_{0,v_0}=\mathbb{R}$ and $K_v=\mathbb{C}$.

In practice, it is however often more practical to use $\|\cdot\|_v$ than $|\cdot|_v$.

Lemma 1.1.10. *Under the assumptions and notation above, we have*

$$\sum_{v|v_0}\log\|x\|_v=[K:K_0]\log|x|_{v_0}\quad\text{for all } x\in K_0^*,$$
$$\sum_{v|v_0}\log\|y\|_v=\log|N_{K/K_0}(y)|_{v_0}\quad\text{for all } y\in K^*.$$

Example 1.1.11. *Again, let us look at the case of number fields. When* $K=\mathbb{Q}$, *set*

$$M_{\mathbb{Q}}:=\{|\cdot|_p : p \text{ prime number or } p=\infty\}$$

normalized as in Example 1.1.2.(ii).

In general for an arbitrary number field K.

- *Each place* v *of* K *lying over* p *corresponds to a unique prime ideal* $\mathfrak{p}$ *dividing* p.

  For each $x\in K^*$, *the fractional ideal* $x\mathcal{O}_K$ *can be uniquely factorized into a finite product* $\prod_{\mathfrak{p}}\mathfrak{p}^{\mathrm{ord}_{\mathfrak{p}}(x)}$ *with* $\mathfrak{p}$ *running over all the prime ideals of* $\mathcal{O}_K$. *This defines a homomorphism* $\mathrm{ord}_{\mathfrak{p}}\colon K^*\to\mathbb{Z}$ *for each* $\mathfrak{p}$ *(and set* $\mathrm{ord}_{\mathfrak{p}}(0):=+\infty$*).*

  Set $|x|_{\mathfrak{p}}:=p^{-[k(\mathfrak{p}):\mathbb{F}_p]\mathrm{ord}_{\mathfrak{p}}(x)/[K_v:\mathbb{Q}_p]}=p^{-\mathrm{ord}_{\mathfrak{p}}(x)/e_{\mathfrak{p}}}$.[3] *Notice that* $|p|_{\mathfrak{p}}=p^{-1}$.

  We show that $|\cdot|_{\mathfrak{p}}$ *is precisely the absolute value* $|\cdot|_v$ *from* (1.1.1). *Indeed, we have for* (1.1.2)

  $$\|x\|_v=|N_{K_{\mathfrak{p}}/\mathbb{Q}_p}(x)|_p=|N_{K_{\mathfrak{p}}/\mathbb{Q}_p}\mathfrak{p}^{\mathrm{ord}_{\mathfrak{p}}(x)}|_p=|p^{[k(\mathfrak{p}):\mathbb{F}_p]\mathrm{ord}_{\mathfrak{p}}(x)}|_p=p^{-[k(\mathfrak{p}):\mathbb{F}_p]\mathrm{ord}_{\mathfrak{p}}(x)}.$$

  It is a standard fact from Algebraic Number Theory that $[K_{\mathfrak{p}}:\mathbb{Q}_p]=e_{\mathfrak{p}}[k(\mathfrak{p}):\mathbb{F}_p]$. *Thus* $|x|_v=\|x\|_v^{1/[K_{\mathfrak{p}}:\mathbb{Q}_p]}$ *equals* $|x|_{\mathfrak{p}}$ *defined above.*

  Now we set

  $$M_K^0:=\{|\cdot|_{\mathfrak{p}} : \mathfrak{p} \text{ prime ideal of } \mathcal{O}_K\}. \tag{1.1.3}$$

  Then each element M_K^0 *is a representative of a non-archimedean place of* K, *and all non-archimedean places of* K *arises in this way.*

- *For an archimedean place* v *of* K, *it lies over the unique archimedean place of* $\mathbb{Q}$ *which gives rise to the embedding* $\mathbb{Q}\hookrightarrow\mathbb{R}$. *Consider all the embeddings* $\sigma\colon K\hookrightarrow\mathbb{C}$; *there are exactly* $[K:\mathbb{Q}]$ *of them. Each such embedding defines an absolute value on* K

  $$|x|_\sigma:=|\sigma(x)|_\infty$$

  where $|z|_\infty$ *is the usual absolute value on* $\mathbb{R}$ *or* $\mathbb{C}$. *It can be shown that all archimedean places of* K *arise in this way.*

  Among the embeddings $K\hookrightarrow\mathbb{C}$ *there are two kinds:* r_1 *real embeddings with* $\sigma(K)\subseteq\mathbb{R}$ *(call them* $\rho_1,\dots,\rho_{r_1}$*) and* r_2 *complex embeddings with* $\sigma(K)\not\subseteq\mathbb{R}$ *(call them* $\tau_1,\overline{\tau}_1,\dots,\tau_{r_2},\overline{\tau}_{r_2}$*). The complex embeddings come in pairs under the complex conjugation. We have* $[K:\mathbb{Q}]=r_1+2r_2$. *One can show that two embeddings* $K\hookrightarrow\mathbb{C}$ *give rise to equivalent absolute values if and only if they are conjugate.*

  In summary, there are r_1+r_2 *archimedean places of* K. *Set*

  $$M_K^\infty:=\{|\cdot|_\sigma\}_{\sigma\in\{\rho_1,\dots,\rho_{r_1},\tau_1,\dots,\tau_{r_2}\}}. \tag{1.1.4}$$

[3] Recall the standard definition $e_{\mathfrak{p}}=\mathrm{ord}_{\mathfrak{p}}(p)$ from Algebraic Number Theory.

Now set

$$M_K = M_K^0 \bigcup M_K^\infty. \tag{1.1.5}$$

From now on, for a number field K we will always use M_K to denote the set from (1.1.5). Moreover, the following convention on the notation M_K for a number field K will always be used in this course.

Notation 1.1.12. *It is sometimes more convenient to work with $\|\cdot\|_v$ than $|\cdot|_v$, and so we will also use the following notation. By $v \in M_K$ for a number field K, we always use $|\cdot|_v$ to denote the corresponding absolute value in the set from (1.1.5), and use $\|\cdot\|_v$ to denote $|\cdot|_v^{[K_v:\mathbb{Q}_p]}$ for $v|p$ and $|\cdot|_v^{[K_v:\mathbb{R}]}$ for $v|\infty$. Notice that when $K = \mathbb{Q}$, $\|\cdot\|_v$ and $|\cdot|_v$ coincide.*

We finish this subsection by the following Product Formula.

Theorem 1.1.13 (Product Formula)**.** *Let K be a number field. Then*

$$\sum_{v \in M_K} \log \|x\|_v = 0 \quad \text{for each } x \in K^*.$$

Proof. Let $x \in K^*$. We start with the case $K = \mathbb{Q}$. In this case, $x = \prod_p p^{\mathrm{ord}_p(x)}$ with p running over all prime numbers. Then

$$\prod_{v \in M_\mathbb{Q}} |x|_v = |x|_\infty \prod_p |x|_p = |x|_\infty \prod_p p^{-\mathrm{ord}_p(x)} = 1.$$

So $\sum_{v \in M_\mathbb{Q}} \log |x|_v = 0$.

For arbitrary K, apply Lemma 1.1.10 to $K/\mathbb{Q}$ and $v|v_0$ with v_0 a place of $M_\mathbb{Q}$. Then we obtain $\sum_{v|p} \log \|x\|_v = \frac{1}{[K:\mathbb{Q}]} \log |N_{K/\mathbb{Q}}(x)|_{v_0}$. So

$$\sum_{v \in M_K} \log \|x\|_v = \sum_{v_0 \in M_\mathbb{Q}} \sum_{v|v_0} \log \|x\|_v = \sum_{v_0 \in M_\mathbb{Q}} \log |N_{K/\mathbb{Q}}(x)|_{v_0},$$

which equals 0 from the case $K = \mathbb{Q}$. Hence we are done. □

1.2 Height on projective spaces

In the whole section, we will use K to denote a number field.

1.2.1 Definition and basic properties

Let us start with the simplest case. Let $x \in \mathbb{P}^1(\mathbb{Q})$. There is a unique way to write x as $[a : b]$ with $a, b \in \mathbb{Z}$ such that we are in one of the following two cases:

- $a = 0$, $b = 1$ or $a = 1$, $b = 0$;
- $a > 0$ and $b \neq 0$ are coprime.

Set

$$H(x) := \max\{|a|, |b|\}.$$

Notice that $H(x) \geq 1$ by definition. Also notice that any rational number x can be identified with $[x : 1] \in \mathbb{P}^1(\mathbb{Q})$, so we can set $H(x) := H([x : 1])$.

In Height Theory, it turns out to be more convenient to work with the *logarithmic height*. On $\mathbb{P}^1(\mathbb{Q})$ it is $h(x) := \log H(x) = \log \max\{|a|, |b|\}$. Then we have $h(x) \geq 0$ for each $x \in \mathbb{P}^1(\mathbb{Q})$.

For more general number fields, we will use the absolute values introduced in the previous section (Example 1.1.11) to define the height. Let $\overline{\mathbb{Q}}$ be an algebraic closure of $\mathbb{Q}$.

Definition 1.2.1. *Let $\mathbf{x} = [x_0 : \cdots : x_n] \in \mathbb{P}^n(\overline{\mathbb{Q}})$. The **(absolute logarithmic Weil) height** of x is defined to be*

$$h(\mathbf{x}) := \frac{1}{[K : \mathbb{Q}]} \sum_{v \in M_K} \log \max\{\|x_0\|_v, \ldots, \|x_n\|_v\},$$

where $K \subseteq \overline{\mathbb{Q}}$ is a number field such that $x_j \in K$ for all j.

*We also set $H(\mathbf{x}) := e^{h(\mathbf{x})}$ to be the **multiplicative height**.*

One can check that this definition coincides with the one for $\mathbb{P}^1(\mathbb{Q})$ above. More generally, we have the following lemma.

Lemma 1.2.2. *Let $\mathbf{x} = [x_0 : \cdots : x_n] \in \mathbb{P}^n(\mathbb{Q})$. Suppose the x_j's are all integers and are coprime. Then*

$$h(\mathbf{x}) = \log \max\{|x_0|, \ldots, |x_n|\}$$

with the usual absolute value.

Proof. Exercise class. □

Lemma 1.2.3. *The height function defined above satisfies the following properties.*

(i) It is independent of the choice of K.

(ii) It is independent of the choice of the homogeneous coordinates.

(iii) $h(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{P}^n(\overline{\mathbb{Q}})$.

Proof. Let $\mathbf{x} = [x_0 : \cdots : x_n] \in \mathbb{P}^n(\overline{\mathbb{Q}})$.

For (i): Assume that each x_j is in K and L for two number fields $K, L \subseteq \overline{\mathbb{Q}}$. We may assume $K \subseteq L$. Then

$$\begin{aligned}
\sum_{w \in M_L} \log \max_j \|x_j\|_w &= \sum_{v \in M_K} \sum_{w|v} \log \max_j \|x_j\|_w \\
&= \sum_{v \in M_K} \sum_{w|v} \log \max_j \|N_{L_w/K_v}(x_j)\|_v \\
&= \sum_{v \in M_K} \sum_{w|v} \log \max_j \|x_j\|_v^{[L_w : K_v]} \\
&= \sum_{v \in M_K} \sum_{w|v} [L_w : K_v] \log \max_j \|x_j\|_v \\
&= \sum_{v \in M_K} [L : K] \log \max_j \|x_j\|_v \quad \text{by Lemma 1.1.8.}
\end{aligned}$$

This establishes (i).

For (ii): Let $[x_0 : \cdots : x_n]$ and $[y_0 : \cdots : y_n]$ be two homogeneous coordinates for a point $\mathbf{x} \in \mathbb{P}^n(\overline{\mathbb{Q}})$. By part (i), we may and do assume that all coordinates are in the same number field K. Then there exists $\lambda \in K^*$ such that $y_j = \lambda x_j$ for each $j \in \{0, \ldots, n\}$. We have then

$$\sum_{v \in M_K} \log \max_j \|y_j\|_v = \sum_{v \in M_K} \log \max_j \|x_j\|_v + \sum_{v \in M_K} \log \|\lambda\|_v = \sum_{v \in M_K} \log \max_j \|x_j\|_v,$$

where the last equality follows from the Product Formula (Theorem 1.1.13). This establishes (ii).

Part (iii) follows from part (ii) because we can always choose homogeneous coordinates for x such that some coordinate is 1. □

Lemma 1.2.4. *The action of the Galois group* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *on* $\mathbb{P}^n(\overline{\mathbb{Q}})$ *leaves the height invariant. More precisely, for any* $\mathbf{x} \in \mathbb{P}^n(\overline{\mathbb{Q}})$ *and any* $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$*, we have* $h(\sigma(\mathbf{x})) = h(\mathbf{x})$.

Proof. Exercise class. □

The following theorem is of fundamental importance for the Height Machine.

Theorem 1.2.5 (Northcott Property)**.** *For each* $B \geq 0$ *and* $D \geq 1$*, the set*

$$\{\mathbf{x} \in \mathbb{P}^n(\overline{\mathbb{Q}}) : h(\mathbf{x}) \leq B,\ [\mathbb{Q}(\mathbf{x}) : \mathbb{Q}] \leq D\}$$

is a finite set.

Proof. We start with the case $D = 1$. Then the set in question becomes

$$\{\mathbf{x} \in \mathbb{P}^n(\mathbb{Q}) : h(\mathbf{x}) \leq B\}.$$

It is not hard to check that this set is finite by Lemma 1.2.2.

For general D. Write $\mathbf{x} = [x_0 : \cdots : x_n] \in \mathbb{P}^n(K)$ such that at least one coordinate equals 1. Then for each $v \in M_K$, we have

$$\max\{\|x_0\|_v, \ldots, \|x_n\|_v\} \geq \max\{\|x_i\|_v, 1\}$$

for each $i \in \{0, \ldots, n\}$. So $B \geq h(\mathbf{x}) \geq h(x_i)$ for each $i \in \{0, \ldots, n\}$. Moreover, $x_i \in K$, and hence $\mathbb{Q}(x_i) \subseteq \mathbb{Q}(x)$ and hence $[\mathbb{Q}(x_i) : \mathbb{Q}] \leq [\mathbb{Q}(x) : \mathbb{Q}] \leq D$.

It suffices to prove that there are finitely many choices for x_i for each $i \in \{0, \ldots, n\}$. Thus it suffices to establish the following simpler finiteness result.
Claim: For each numbers $B \geq 0$ and $d \geq 1$, the set

$$\{x \in \overline{\mathbb{Q}} : h(x) \leq B,\ [\mathbb{Q}(x) : \mathbb{Q}] = d\}$$

is finite.

Let us prove this claim. Write $K = \mathbb{Q}(x)$, and write $x_1 = x, \ldots, x_d$ for the Galois conjugates of x over $\mathbb{Q}$. The minimal polynomial of x over $\mathbb{Q}$ is

$$F(T) = \prod_{j=1}^{d} (T - x_j) = \sum_{r=0}^{d} (-1)^r s_r(x) T^{d-r}$$

with $s_r(x)$ the r-th symmetric polynomial in $x_1, \dots, x_d$. Denote by $s_r = s_r(x)$; it is a number in $\mathbb{Q}$. For each $v \in M_K$ we have

$$\begin{aligned} |s_r|_v &= \left| \sum_{1 \le i_1 < \cdots < i_r \le d} x_{i_1} \cdots x_{i_r} \right|_v \\ &\le \epsilon(v,r,d) \max_{1 \le i_1 < \cdots < i_r \le d} |x_{i_1} \cdots x_{i_r}|_v \quad \text{triangular inequality} \\ &\le \epsilon(v,r,d) \max_{1 \le i \le d} |x_i|_v^r. \end{aligned}$$

Here one can take $\epsilon(v,r,d) = 1$ if v is non-archimedean and $\epsilon(v,r,d) = \binom{d}{r} \le 2^d$ if v is archimedean.

Thus we have $\|s_r\|_v \le \max_{1 \le i \le d} \|x_i\|_v^r$ if v is non-archimedean, and $\|s_r\|_v \le 2^{d[K_v:\mathbb{R}]} \max_{1 \le i \le d} \|x_i\|_v^r$ if v is archimedean.

Consider the point $s := [s_0 : \cdots : s_d : 1] \in \mathbb{P}^{d+1}(\mathbb{Q})$. We have

$$\begin{aligned} [K:\mathbb{Q}]h(s) &= \sum_{v \in M_K} \log \max_{0 \le r \le d} \{\|s_r\|_v, 1\} \\ &= \sum_{v \in M_K} \max_{0 \le r \le d} \{\log \|s_r\|_v, 0\} \\ &\le \sum_{v \in M_K} \max_{0 \le r \le d} \max_{1 \le i \le d} \{r \log \|x_i\|_v, 0\} + d \sum_{v \mid \infty} [K_v : \mathbb{R}] \log 2 \\ &\le \sum_{v \in M_K} d \max_{1 \le i \le d} \{\log \|x_i\|_v, 0\} + d \sum_{v \mid \infty} [K_v : \mathbb{R}] \log 2 \\ &\le d \sum_{1 \le i \le d} \sum_{v \in M_K} \max\{\log \|x_i\|_v, 0\} + d \sum_{v \mid \infty} [K_v : \mathbb{R}] \log 2 \\ &= d \sum_{1 \le i \le d} [K:\mathbb{Q}] h(x_i) + d \sum_{v \mid \infty} [K_v : \mathbb{R}] \log 2 \\ &= d[K:\mathbb{Q}] \cdot dh(x) + d[K:\mathbb{Q}] \log 2 \quad \text{by Lemma 1.2.4.} \end{aligned}$$

So $h(s) \le d^2 h(x) + d \log 2 \le d^2 B + d \log 2$ is bounded. But $s \in \mathbb{P}^{d+1}(\mathbb{Q})$, so by the case $D = 1$ there are only finitely many choices for s. So there are only finitely many choices for $s_0, \dots, s_d$, and therefore only finitely many choices for the minimal polynomial of x over $\mathbb{Q}$. Thus there are only finitely many choices for x, and this is exactly the desired claim. We are done. □

1.2.2 Height on affine spaces

In the proof of the Northcott property, we computed the height of $[s_0 : \cdots : s_d : 1] \in \mathbb{P}^{d+1}(\overline{\mathbb{Q}})$. This point lies in $\mathbb{A}^{d+1}(\overline{\mathbb{Q}})$, viewed as the complement of the hypersurface with last homogeneous coordinate being 0. It is then convenient to introduce the following notions.

Notation 1.2.6. *Set*

$$\log^+(x) := \max\{\log x, 0\} = \log \max\{x, 1\}$$

for each $x > 0$.

Definition 1.2.7. *For each point* $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{A}^n(\overline{\mathbb{Q}})$*, define*

$$h(\mathbf{x}) := h([\mathbf{x} : 1])$$

with $[\mathbf{x}:1] := [x_1 : \cdots : x_n : 1]$ *viewed as a point in* $\mathbb{P}^n(\overline{\mathbb{Q}})$.
We also set $H(\mathbf{x}) := e^{h(\mathbf{x})}$.

We have

$$h(\mathbf{x}) = \max_{1\le j\le n}\{h(x_j)\} = \frac{1}{[K:\mathbb{Q}]}\sum_{v\in M_K}\log^+\max\{\|x_1\|_v,\dots,\|x_n\|_v\}. \tag{1.2.1}$$

The next proposition discusses the height of the sum of algebraic numbers. It will be seen again in the discussion for heights of polynomials.

Proposition 1.2.8. *Let* $P_1,\dots,P_r \in \mathbb{A}^n(\overline{\mathbb{Q}})$. *Then*

$$h(P_1+\cdots+P_r) \le h(P_1)+\cdots+h(P_r)+\log r.$$

In the case $n=1$, the left hand side is the sum of r algebraic numbers.

Proof. Write, for each $k\in\{1,\dots,r\}$, $P_k = (x_1^{(k)},\dots,x_n^{(k)})$. Assume all the P_k's are in a number field K. Then

$$[K:\mathbb{Q}]h(P_1+\cdots+P_r) = \sum_{v\in M_K}\max_{1\le j\le n}\log^+\|x_j^{(1)}+\cdots+x_j^{(r)}\|_v.$$

If v is not archimedean, then $\|\cdot\|_v$ is an absolute value and hence

$$\|x_j^{(1)}+\cdots+x_j^{(r)}\|_v \le \max_{1\le k\le r}\|x_j^{(k)}\|_v.$$

If v is archimedean, then the triangular inequality for the absolute value $|\cdot|_v$ yields $|x_j^{(1)}+\cdots+x_j^{(r)}|_v \le |r|_v\max_{1\le k\le r}|x_j^{(k)}|_v$. Hence raising both sides to the power of $[K_v:\mathbb{R}]$ we get

$$\|x_j^{(1)}+\cdots+x_j^{(r)}\|_v \le \|r\|_v\max_{1\le k\le r}\|x_j^{(k)}\|_v$$

Thus

$$\begin{aligned}
[K:\mathbb{Q}]h(P_1+\cdots+P_r) &\le \sum_{v\in M_K}\max_{j,k}\log^+\|x_j^{(k)}\|_v + \sum_{v|\infty}\log\|r\|_v\\
&\le \sum_{1\le k\le r}\sum_{v\in M_K}\max_{j}\log^+\|x_j^{(k)}\|_v + \sum_{v|\infty}\log\|r\|_v\\
&= \sum_{1\le k\le r}[K:\mathbb{Q}]h(P_k) + [K:\mathbb{Q}]\log r \quad \text{by Lemma 1.1.10.}
\end{aligned}$$

Hence we are done. □

1.2.3 Liouville's inequality

Lemma 1.2.9. $h(1/\alpha) = h(\alpha)$ *for any* $\alpha\in K^*$.

Proof. By definition, $h(1/\alpha) = h([1/\alpha:1])$ with $[1/\alpha:1]\in\mathbb{P}^1(K)$ and similarly $h(\alpha)=h([\alpha:1])$. So

$$h(1/\alpha) = h([1/\alpha:1]) = h([1:\alpha]) = h([\alpha:1]) = h(\alpha),$$

with $h([1:\alpha]) = h([\alpha:1])$ following directly from the definition of height. □

Alternatively, one can check

$$\log|\alpha|_v = \log^+|\alpha|_v - \log^+|1/\alpha|_v \tag{1.2.2}$$

and use the Product Formula to prove this lemma.

Proposition 1.2.10 (Fundamental Inequality)**.** *Let $S \subseteq M_K$ be a finite set. For each $\alpha \in K^*$, we have*

$$h(\alpha) \geq \frac{1}{[K:\mathbb{Q}]} \sum_{v\in S} \log\|\alpha\|_v \tag{1.2.3}$$

and

$$h(\alpha) \geq -\frac{1}{[K:\mathbb{Q}]} \sum_{v\in S} \log\|\alpha\|_v. \tag{1.2.4}$$

Proof. By the definition $h(\alpha) = \frac{1}{[K:\mathbb{Q}]}\sum_{v\in M_K}\log^+\|\alpha\|_v$ and noticing that $\log^+$ takes non-negative values, we get the first inequality.

To prove second inequality, we apply the first inequality to $1/\alpha$ and use Lemma 1.2.9. □

Example 1.2.11. *Consider $K = \mathbb{Q}$ and $\alpha = p$ is a prime number. Then $h(p) = \log p$, $|p|_\infty = p$ and $|p|_p = p^{-1}$. Now* (1.2.3) *attains equality for $S = \{\infty\}$, and* (1.2.4) *attains equality for $S = \{p\}$.*

Now we are ready to state Liouville's inequality. The classical formulation is in terms of the multiplicative height $H(\cdot) = e^{h(\cdot)}$.

In the statement of Liouville's Inequality, let K_0 be a number field.

Theorem 1.2.12 (Liouville's Inequality)**.** *Fix $\beta \in K_0$. Let K/K_0 be a finite extension and consider a finite set $S \subseteq M_K$. For any $\alpha \in K$ with $\alpha \neq \beta$, we have*

$$\prod_{v\in S}\|\alpha-\beta\|_{v,K_0} \geq (2H(\alpha)H(\beta))^{-[K:K_0]},$$

where $\|\cdot\|_{v,K_0} := \|\cdot\|_v^{1/[K_0:\mathbb{Q}]}$.

Before moving on to its proof, let us look at the following corollary which is closer to the classical statement of this inequality. It has a flavor of approximating algebraic numbers by rational numbers.

Corollary 1.2.13. *Let $\alpha \in \mathbb{R}$ be an algebraic number of degree $r > 1$, i.e. $[\mathbb{Q}(\alpha):\mathbb{Q}] = r$. Then there exists a constant $c(\alpha) > 0$ such that for the usual absolute value $|\cdot|$ on $\mathbb{R}$, we have*

$$|\alpha-\beta| \geq c(\alpha)H(\beta)^{-r} \quad \text{for all } \beta \in \mathbb{Q}.$$

This corollary follows immediately from Theorem 1.2.12 applied to $K_0 = \mathbb{Q}$, $K = \mathbb{Q}(\alpha)$ and S the archimedean place given by the natural inclusion $\mathbb{Q}(\alpha) \subseteq \mathbb{R}$. Notice that if $\alpha \in \mathbb{C}\setminus\mathbb{R}$, then the same conclusion holds true with $|\cdot|$ replaced by $|\cdot|^2$.

Proof of Theorem 1.2.12. Apply Proposition 1.2.8 to $n = 1$, $r = 2$, $P_1 = \alpha$ and $P_2 = -\beta$. Then we get $h(\alpha-\beta) \leq h(\alpha) + h(\beta) + \log 2$. So $H(\alpha-\beta) \leq 2H(\alpha)H(\beta)$.

Apply the Fundamental Inequality (1.2.4) to $\alpha - \beta$. Then we get

$$h(\alpha-\beta) \geq -\frac{1}{[K:\mathbb{Q}]}\sum_{v\in S}\log\|\alpha-\beta\|_v = \frac{1}{[K:\mathbb{Q}]}\log(\prod_{v\in S}\|\alpha-\beta\|_v)^{-1}.$$

From this, we get

$$\prod_{v\in S}\|\alpha-\beta\|_{v,K_0}=(\prod_{v\in S}\|\alpha-\beta\|_v)^{1/[K_0:\mathbb{Q}]}\geq(e^{h(\alpha-\beta)})^{-[K:K_0]}=H(\alpha-\beta)^{-[K:K_0]}.$$

Now we can conclude because we have seen $H(\alpha-\beta)\leq 2H(\alpha)H(\beta)$. □

1.2.4 The change of height under geometric operations

In this section, we go back to the height function on $\mathbb{P}^n(\overline{\mathbb{Q}})$. We will consider several geometric operations concerning projective spaces and see how the heights change.

Consider the *Segre embedding*

$$S_{n,m}\colon \mathbb{P}^n\times\mathbb{P}^m\to\mathbb{P}^{(n+1)(m+1)-1},\quad (\mathbf{x},\mathbf{y})\mapsto\mathbf{x}\otimes\mathbf{y}:=(x_iy_j)_{i,j} \tag{1.2.5}$$

and the *d-uple embedding*

$$\Phi_d\colon \mathbb{P}^n\to\mathbb{P}^N,\quad \mathbf{x}\mapsto[M_0(\mathbf{x}):\cdots:M_N(\mathbf{x})] \tag{1.2.6}$$

with $N=\binom{n+d}{n}-1$ and $\{M_0(\mathbf{x}),\dots,M_N(\mathbf{x})\}$ the complete collection of monomials of degree d in the variables $x_0,\dots,x_n$.

Proposition 1.2.14. *We have*

(i) $h(\mathbf{x}\otimes\mathbf{y})=h(\mathbf{x})+h(\mathbf{y})$ *for all* $\mathbf{x}\in\mathbb{P}^n(\overline{\mathbb{Q}})$ *and* $\mathbf{y}\in\mathbb{P}^m(\overline{\mathbb{Q}})$.

(ii) $h(\Phi_d(\mathbf{x}))=dh(\mathbf{x})$ *for all* $\mathbf{x}\in\mathbb{P}^n(\overline{\mathbb{Q}})$.

Proof. Part (i) in Exercise class, by using $\max_{i,j}|x_iy_j|_v=\max_i|x_i|_v\cdot\max_j|y_j|_v$.

We prove part (ii). Each $M_i(\mathbf{x})$ is a monomial of degree d in the variables $x_0,\dots,x_n$. It is clear that $|M_j(\mathbf{x})|_v\leq\max_i|x_i|_v^d$ for each $0\leq j\leq N$. Moreover since the particular monomials $x_0^d,\dots,x_n^d$ appear in the collection, we have

$$\max_{0\leq j\leq N}|M_j(\mathbf{x})|_v=\max_{0\leq i\leq n}|x_i|_v^d.$$

From this we can conclude. □

We finish this section by a discussion on the change of heights under linear maps.

Theorem 1.2.15. *Let* $\phi\colon\mathbb{P}^n\to\mathbb{P}^m$ *be a linear map defined over* $\overline{\mathbb{Q}}$*, i.e.* $\phi=[L_0(\mathbf{x}):\cdots:L_m(\mathbf{x})]$ *for some linear forms on* $\mathbb{P}^n$*. Let* $Z\subseteq\mathbb{P}^n$ *be the common zero of the* L_i*'s.*

Let $X\subseteq\mathbb{P}^n$ *be a closed subvariety such that* $X\cap Z=\emptyset$*. Then*

$$h(\phi(\mathbf{x}))=h(\mathbf{x})+O(1)\quad \text{for all } \mathbf{x}\in X(\overline{\mathbb{Q}}).$$

More precisely, the conclusion means that there exists a constant $c=c(\phi,X)>0$ depending only on ϕ and X such that

$$|h(\phi(\mathbf{x}))-h(\mathbf{x})|\leq c$$

for all $\mathbf{x}\in X(\overline{\mathbb{Q}})$. We remark that this bound[4] does not hold true on the whole $\mathbb{P}^n\setminus Z$, but on any closed subvariety disjoint from Z.

[4] Or more precisely, "half" of the bound does not hold true on the whole $\mathbb{P}^n\setminus Z$ as will be shown in the proof.

Proof. The proof is divided into two parts. We may and do assume that $Z \neq \mathbb{P}^n$, *i.e.* one of the L_i's does not vanish on the whole $\mathbb{P}^n$.

Write $L_i(\mathbf{x}) = \sum_{0\leq j\leq n} a_{i,j}x_j$. Then $[a_{0,0} : \cdots : a_{0,n} : \cdots : a_{m,0} : \cdots : a_{m,n}]$ is a point in $\mathbb{P}^{(n+1)(m+1)-1}(\overline{\mathbb{Q}})$ and is uniquely determined by ϕ.

Part I Prove: there exists a constant $c_1(\phi)$ depending only on ϕ such that $h(\phi(\mathbf{x}))-h(\mathbf{x}) \leq c_1(\phi)$ for all $\mathbf{x} \in (\mathbb{P}^n \setminus Z)(\overline{\mathbb{Q}})$.

Let $\mathbf{x} = [x_0 : \cdots : x_n] \in (\mathbb{P}^n \setminus Z)(\overline{\mathbb{Q}})$. Fix a number field K such that $\mathbf{x} \in \mathbb{P}^n(K)$ and all $a_{i,j}$'s are in K. Then for each $v \in M_K$, we have

$$|L_i(\mathbf{x})|_v = |\sum_{0\leq j\leq n} a_{i,j}x_j|_v \leq \epsilon(v, n+1)(\max_j |a_{i,j}|_v)(\max_j |x_j|_v)$$

where $\epsilon(v,k) := \begin{cases} 1 & \text{if } v \text{ is non-archimedean} \\ k & \text{if } v \text{ is archimedean} \end{cases}$. Raising both sides to the power of $[K_v : \mathbb{Q}_p]$ (with $\mathbb{Q}_\infty = \mathbb{R}$), we get

$$\max_i \|L_i(\mathbf{x})\|_v \leq \epsilon(v, n+1)^{[K_v:\mathbb{Q}_p]}(\max_{i,j} \|a_{i,j}\|_v)(\max_j \|x_j\|_v).$$

Now we have

$$\begin{aligned}
[K : \mathbb{Q}]h(\phi(\mathbf{x})) &= \sum_{v\in M_K} \log \max_i \|L_i(\mathbf{x})\|_v \\
&\leq \sum_{v\in M_K} \log\left(\epsilon(v, n+1) \cdot \max_{i,j} \|a_{i,j}\|_v \cdot \max_j \|x_j\|_v\right) \\
&\leq \sum_{v\in M_K} (\log \max_{i,j} \|a_{i,j}\|_v + \log \max_j \|x_j\|_v) + \sum_{v|\infty} [K_v : \mathbb{R}] \log(n+1) \\
&= \sum_{v\in M_K} \log \max_{i,j} \|a_{i,j}\|_v + [K : \mathbb{Q}]h(\mathbf{x}) + [K : \mathbb{Q}] \log(n+1) \\
&= [K : \mathbb{Q}]h([a_{0,0} : \cdots : a_{0,n} : \cdots : a_{m,0} : \cdots : a_{m,n}]) + [K : \mathbb{Q}]h(\mathbf{x}) + [K : \mathbb{Q}] \log(n+1).
\end{aligned}$$

Thus $h(\phi(\mathbf{x})) - h(\mathbf{x}) \leq h([a_{0,0} : \cdots : a_{0,n} : \cdots : a_{m,0} : \cdots : a_{m,n}]) + \log(n+1)$. The first term on the right hand side depends only on ϕ. So we are done for this part.

Part II Prove: there exists a constant constant $c_2(\phi, X)$ such that $h(\phi(\mathbf{x})) - h(\mathbf{x}) \geq c_2(\phi, X)$ for all $\mathbf{x} \in X(\overline{\mathbb{Q}})$.

Write $I(X) = (F_1, \ldots, F_r)$. Since $X \cap Z = \emptyset$, we have that the polynomials $L_0, \ldots, L_m, F_1, \ldots, F_r$ have no common zeros in $\mathbb{P}^n$. By Hilbert Nullstellensatz, we then have the following equality of ideals of $\overline{\mathbb{Q}}[X_0, \ldots, X_n]$

$$\sqrt{(L_0, \ldots, L_m, F_1, \ldots, F_r)} = (X_0, \ldots, X_n).$$

In particular, for each $j \in \{0, \ldots, n\}$, we can find polynomials $G_{i,j}$ and $H_{i,j}$ and an exponent $t \geq 1$, all depending only on X and ϕ, such that

$$G_{0,j}L_0 + \cdots + G_{m,j}L_m + H_{1,j}F_1 + \cdots + H_{r,j}F_r = X_j^t.$$

Moreover $\deg G_{i,j} = t - \deg L_i = t - 1$.

Write $G_{i,j} = \sum_{|\mathbf{e}|=t-1} b_{i,j,\mathbf{e}}X^{\mathbf{e}}$, with $\mathbf{e} = (e_0, \ldots, e_n)$ a multi-index with $|\mathbf{e}| := e_0 + \cdots + e_n$ and $X^{\mathbf{e}} = X_0^{e_0} \cdots X_n^{e_n}$. Notice that $G_{i,j}$ is the sum of at most $\binom{n+t-1}{n}$ monomials.

Now let $\mathbf{x} = [x_0 : \cdots : x_n] \in X(\overline{\mathbb{Q}})$. Evaluating the equation above at $\mathbf{x}$, we get

$$G_{0,j}(\mathbf{x})L_0(\mathbf{x}) + \cdots + G_{m,j}(\mathbf{x})L_m(\mathbf{x}) = x_j^t$$

for each $j \in \{0, \ldots, n\}$.

Fix a number field K such that $\mathbf{x} \in X(K)$ and all the coefficients $b_{i,j,\mathbf{e}}$ are in K.

For each $v \in M_K$, we have

$$|x_j|_v^t = |G_{0,j}(\mathbf{x})L_0(\mathbf{x}) + \cdots + G_{m,j}(\mathbf{x})L_m(\mathbf{x})|_v \leq \epsilon(v, m+1) \max_{0\leq i\leq m} |G_{i,j}(\mathbf{x})|_v \max_{0\leq i\leq m} |L_i(\mathbf{x})|_v.$$

Thus

$$\begin{aligned}
\max_j |x_j|_v^t &\leq \epsilon(v, m+1) \max_{i,j} |G_{i,j}(\mathbf{x})|_v \max_i |L_i(\mathbf{x})|_v \\
&= \epsilon(v, m+1)(\max_{i,j} |\sum_{|\mathbf{e}|=t-1} b_{i,j,\mathbf{e}} x_0^{e_0} \cdots x_n^{e_n}|_v)(\max_{0\leq i\leq m} |L_i(\mathbf{x})|_v) \\
&\leq \epsilon(v, m+1) \left(\epsilon(v, \binom{n+t-1}{n}) \max_{i,j,\mathbf{e}} |b_{i,j,\mathbf{e}}|_v \max_j |x_j|_v^{t-1}\right) (\max_{0\leq i\leq m} |L_i(\mathbf{x})|_v).
\end{aligned}$$

Dividing both sides by $\max_j |x_j|^{t-1}$, we get

$$\max_j |x_j|_v \leq \epsilon(v, m+1)\epsilon(v, \binom{n+t-1}{n}) \max_{i,j,\mathbf{e}} |b_{i,j,\mathbf{e}}|_v \max_{0\leq i\leq m} |L_i(\mathbf{x})|_v.$$

Raising both sides to the power of $[K_v : \mathbb{Q}_p]$ (with $\mathbb{Q}_\infty = \mathbb{R}$), we get

$$\max_j \|x_j\|_v \leq \epsilon(v, m+1)^{[K_v:\mathbb{Q}_p]}\epsilon(v, \binom{n+t-1}{n})^{[K_v:\mathbb{Q}_p]} \max_{i,j} \|b_{i,j,\mathbf{e}}\|_v \max_i \|L_i(\mathbf{x})\|_v.$$

Now we have

$$\begin{aligned}
[K : \mathbb{Q}]h(\mathbf{x}) &= \sum_{v\in M_K} \max_j \|x_j\|_v \\
&\leq \sum_{v\in M_K} \log \max_{i,j,\mathbf{e}} \|b_{i,j,\mathbf{e}}\|_v + \sum_{v\in M_K} \max_i \|L_i(\mathbf{x})\|_v + \sum_{v|\infty} [K_v : \mathbb{R}] \log(m+1)\binom{n+t-1}{n} \\
&= [K : \mathbb{Q}]h(\mathbf{b}) + [K : \mathbb{Q}]h(\phi(\mathbf{x})) + [K : \mathbb{Q}] \log(m+1)\binom{n+t-1}{n}
\end{aligned}$$

where $\mathbf{b}$ is the point in an appropriate projective space whose homogeneous coordinates are $b_{i,j,\mathbf{e}}$. Notice that $\mathbf{b}$ is uniquely determined by the $G_{i,j}$'s, and hence by X and ϕ. Now we get the desired inequality $h(\phi(\mathbf{x})) - h(\mathbf{x}) \geq c_2(\phi, X)$ for all $\mathbf{x} \in X(\overline{\mathbb{Q}})$. Hence we are done. □

1.3 Height of polynomials

In this section, we study the heights of polynomials. We will use the Weil height on projective and affine spaces defined in §1.2 of this chapter.

Definition 1.3.1. *The (affine) height of a polynomial*

$$f(t_1, \ldots, t_n) = \sum_{j_1,\ldots,j_n} a_{j_1\ldots j_n} t_1^{j_1} \cdots t_n^{j_n} = \sum_{\mathbf{j}} a_{\mathbf{j}} \mathbf{t}^{\mathbf{j}}$$

with coefficients in $\overline{\mathbb{Q}}$ is the quantity $h(\mathbf{a})$ where $\mathbf{a} = (a_{\mathbf{j}})_{\mathbf{j}}$ is viewed as a point in $\overline{\mathbb{Q}}^N$ for some N.

In other words, if we assume each $a_{\mathbf{j}} \in K$ for an appropriate number field K and define the **Gauß norm**

$$\|f\|_v := \max_{\mathbf{j}} \|a_{\mathbf{j}}\|_v \tag{1.3.1}$$

for each $v \in M_K$, then we have

$$h(f) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in M_K} \log^+ \|f\|_v = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in M_K} \log \max\{\|f\|_v, 1\}. \tag{1.3.2}$$

1.3.1 Affine height vs the Projective height

In some literature, one defines the height of f as the height of the point $[a_{\mathbf{j}}]_{\mathbf{j}}$ viewed as a point in an appropriate *projective* space. This is sometimes called the *projective height of* f, and is in general smaller than the affine height we defined above.

In this course, we always use the affine height. An important advantage to take this convention is the following proposition, which is about the evaluation of a polynomial at a point. The proof shares some similarities with the proof of Theorem 1.2.15.

Proposition 1.3.2. *Let d be the sum the partial degrees of f. Let $\mathbf{x} = (x_1, \dots, x_n) \in \overline{\mathbb{Q}}^n$. Then*

$$h(f(\mathbf{x})) \le h(f) + dh(\mathbf{x}) + \min\{(n+1)\log(n+d+1), (n+d+1)\log 2\}.$$

As shown by the proof, this result is not correct if we use the projective height of f.

Proof. Write $f(\mathbf{t}) = \sum_{k=0}^{d} \sum_{|\mathbf{j}|=k} a_{\mathbf{j}} \mathbf{t}^{\mathbf{j}}$. Set $\psi(n,d) := \min\{(n+d+1)^{n+1}, 2^{n+d+1}\}$. Then as in the proof of Theorem 1.2.15, it is not hard to check

$$|\sum_{|\mathbf{j}|=k} a_{\mathbf{j}} \mathbf{x}^{\mathbf{j}}|_v \le \epsilon\left(v, \binom{n+k}{n}\right) \max_{|\mathbf{j}|=k}\{|a_{\mathbf{j}}|_v\} \max_i |x_i|_v^k \le \epsilon\left(v, \binom{n+k}{n}\right) \max_{|\mathbf{j}|=k}\{|a_{\mathbf{j}}|_v\} \max\{1, \max_i |x_i|_v\}^d$$

with $\epsilon(v,m)$ defined to be 1 for v non-archimedean and to be m for v archimedean. Recall that $\sum_{k=0}^{d} \binom{n+k}{n} = \binom{n+d+1}{n+1} \le \psi(n,d)$. So

$$|f(\mathbf{x})|_v = |\sum_{\mathbf{j}} a_{\mathbf{j}} \mathbf{x}^{\mathbf{j}}|_v \le \sum_{i=0}^{k} |\sum_{|\mathbf{j}|=k} a_{\mathbf{j}} \mathbf{x}^{\mathbf{j}}|_v \le \epsilon(v, \psi(n,d)) \max_{\mathbf{j}}\{|a_{\mathbf{j}}|_v\} \max_i\{1, |x_i|_v\}^d$$

and hence

$$\max\{1, |f(\mathbf{x})|_v\} \le \epsilon(v, \psi(n,d)) \max_{\mathbf{j}}\{1, |a_{\mathbf{j}}|_v\} \max_i\{1, |x_i|_v\}^d.$$

Raising to the power of $[K_v : \mathbb{Q}_p]$ and taking the log, we get an upper bound for $\log^+ \|f(\mathbf{x})\|_v$. Hence

$$\begin{aligned}
[K:\mathbb{Q}]h(f(\mathbf{x})) &= \sum_{v \in M_K} \log^+ \|f(\mathbf{x})\|_v \\
&\le \sum_{v \in M_K} \max_{\mathbf{j}} \log^+ \|a_{\mathbf{j}}\|_v + d \sum_{v \in M_K} \max_i \log^+ \|x_i\|_v + \sum_{v|\infty} [K_v : \mathbb{R}] \log \psi(n,d) \\
&= [K:\mathbb{Q}]h(f) + [K:\mathbb{Q}]dh(\mathbf{x}) + [K:\mathbb{Q}] \log \psi(n,d).
\end{aligned}$$

We are done. □

1.3.2 Height of product

The main result of this section is to study the height of a product of two polynomials.

First, an immediate corollary of Proposition 1.2.14.(i) is

Lemma 1.3.3. *Let $f(t_1, \dots, t_n)$ and $g(s_1, \dots, s_m)$ be two polynomials in disjoint sets of variables. Then*

$$h(fg) = h(f) + h(g).$$

However, if f and g do not have disjoint sets of variables, the estimate of $h(fg)$ in terms of $h(f)$ and $h(g)$ is more complicated. We will prove the following theorem for this estimate.

Theorem 1.3.4. *Let $f_1, \dots, f_m$ be polynomials in n variables with coefficients in $\overline{\mathbb{Q}}$. Let d be the sum of the partial degrees of $f := f_1 \cdots f_m$. Then*

$$-d \log 2 + \sum_{j=1}^{m} h(f_j) \leq h(f) \leq d \log 2 + \sum_{j=1}^{m} h(f_j).$$

Moreover in the second inequality, one can replace d by the sum of the partial degrees of the product $f_1 \cdots f_{m-1}$.

To prove this theorem, one separates the non-archimedean places and the archimedean places. For the non-archimedean places, we prove *Gauß's Lemma*. For the archimedean places, we prove *Gelfond's Lemma*. Then we combine these two lemmas to conclude.

Non-archimedean places

The contribution at the non-archimedean places is not hard to study. In this case, we have the following:

Lemma 1.3.5 (Gauß's Lemma)**.** *If v is non-archimedean, then $\|fg\|_v = \|f\|_v \|g\|_v$.*

Proof. The direction $\|fg\|_v \leq \|f\|_v \|g\|_v$ is not hard to obtain because v is non-archimedean.

Now we focus on proving the other direction $\|fg\|_v \geq \|f\|_v \|g\|_v$.

One-variable case We start with the case where both $f(t) = \sum_j a_j t^j$ and $g(t) = \sum_j b_j t^j$ are polynomials in one variable t. Up to dividing both f and g by an appropriate element in K, we may and do assume $\|f\|_v = \|g\|_v = 1$. Then $\|fg\|_v \leq 1$.

Suppose $\|fg\|_v < 1$ and we wish to get a contradiction.

For each j, set $c_j = \sum_{j=k+l} a_k b_l$. Then $fg = \sum_j c_j t^j$. Let j_0 be the smallest integer with $\|a_{j_0}\|_v = 1$. Since $\|a_k\|_v < 1$ for each $k < j_0$, we have $\|a_k b_{j_0-k}\|_v < 1$ for each $k < j_0$. If $\|b_0\|_v = 1$, then $\|a_{j_0} b_0\|_v = 1$ and hence $\|c_{j_0}\|_v = \|a_{j_0} b_0 + \sum_{k<j_0} a_k b_{j_0-k}\|_v = 1$, contradicting $\|fg\|_v < 1$. Hence $\|b_0\|_v < 1$.

Next for each l_0, we prove that $\|b_{l_0}\|_v < 1$ by induction. Suppose we have proved for $0, \dots, l_0 - 1$. Consider $c_{j_0+l_0} = \sum_{0 \leq k \leq j_0+l_0} a_k b_{j_0+l_0-k}$. For $0 \leq k \leq j_0 - 1$, we have $\|a_k\|_v < 1$ and hence $\|a_k b_{j_0+l_0-k}\|_v < 1$. For $j_0 + 1 \leq k \leq j_0 + l_0$, we have $\|b_{j_0+l_0-k}\|_v < 1$ by induction hypothesis and hence $\|a_k b_{j_0+l_0-k}\|_v < 1$. Thus $\|b_{l_0}\|_v = 1$ would yield $\|c_{j_0+l_0}\|_v = \|a_{j_0} b_{l_0}\|_v = 1$, contradicting $\|fg\|_v < 1$. Hence we can conclude $\|b_{l_0}\|_v < 1$.

But then $\|g\|_v < 1$, contradicting $\|g\|_v = 1$. So we can conclude that $\|fg\|_v = 1$ for this case.

General case Write $f(x_1, \dots, x_n) = \sum_{\mathbf{j}} a_{\mathbf{j}} \mathbf{x}^{\mathbf{j}}$ and $g(x_1, \dots, x_n) = \sum_{\mathbf{j}} b_{\mathbf{j}} \mathbf{x}^{\mathbf{j}}$. One can reduce the general case to the one-variable case by the following standard technique. Fix an integer $d > \deg(fg)$, and consider the **Kronecker substitution**

$$x_j := t^{d^{j-1}} \quad (j = 1, \dots, n). \tag{1.3.3}$$

Then

$$f(x_1,\dots,x_n)=\sum_{\mathbf{j}} a_{\mathbf{j}}(t^{d^0})^{j_1}(t^{d^1})^{j_2}\cdots(t^{d^{n-1}})^{j_n}=\sum_{0\le j_1,\dots,j_n\le d-1} a_{j_1,\dots,j_n}t^{j_1+dj_2+\cdots+d^{n-1}j_n},$$
$$\text{and}\quad g(x_1,\dots,x_n)=\sum_{0\le j_1,\dots,j_n\le d-1} b_{j_1,\dots,j_n}t^{j_1+dj_2+\cdots+d^{n-1}j_n}.$$

It is not hard to see that both $f_0(t):=\sum_{0\le j_1,\dots,j_n\le d-1} a_{j_1,\dots,j_n}t^{j_1+dj_2+\cdots+d^{n-1}j_n}$ and $g_0(t):=\sum_{0\le j_1,\dots,j_n\le d-1} b_{j_1,\dots,j_n}t^{j_1+dj_2+\cdots+d^{n-1}j_n}$ are one-variable polynomials in simplified form. So $\|f_0\|_v=\max_{j_1,\dots,j_n}\|a_{j_1,\dots,j_n}\|_v=\|f\|_v$, $\|g_0\|_v=\max_{j_1,\dots,j_n}\|b_{j_1,\dots,j_n}\|_v=\|g\|_v$, and $\|f_0g_0\|_v=\|fg\|_v$. Hence we can conclude by the one-variable case. □

Archimedean places

It is more complicated to handle the archimedean places. The goal is to prove *Gelfond's Lemma* (Lemma 1.3.6), which plays a similar role as Gauß's Lemma for the archimedean places.

In this subsection, we consider polynomials with coefficients in $\mathbb{C}$. We use $|\cdot|$ to denote the usual euclidean absolute value on $\mathbb{C}$.

Let $f=\sum_{\mathbf{j}} a_{\mathbf{j}}\mathbf{t}^{\mathbf{j}}\in\mathbb{C}[t_1,\dots,t_n]$. Define

$$\ell_\infty(f)=|f|_\infty:=\max_{\mathbf{j}}|a_{\mathbf{j}}|. \tag{1.3.4}$$

We also call $\ell_\infty(f)$ the *L^∞-norm* of f.

Now we can state the main result of this subsection.

Lemma 1.3.6 (Gelfond's Lemma)**.** *Let $f_1,\dots,f_m\in\mathbb{C}[t_1,\dots,t_n]$ and set $f:=f_1\cdots f_m$. Let d be the sum of the partial degrees of f. Then*

$$2^{-d}\prod_{j=1}^m\ell_\infty(f_j)\le\ell_\infty(f)\le 2^d\prod_{j=1}^m\ell_\infty(f_j).$$

Moreover in the second inequality, one can replace d by the sum of the partial degrees of the product $f_1\cdots f_{m-1}$.

Before moving on, let us see how Gauß's Lemma and Gelfond's Lemma imply Theorem 1.3.4.

Proof of Theorem 1.3.4. We have

$$[K:\mathbb{Q}]h(f)=\sum_{v\in M_K}\log^+\|f\|_v=\sum_{v\in M_K}\log\max\{\|f\|_v,1\}=\sum_{v\in M_K}\log\max\{\|f_1\cdots f_m\|_v,1\}.$$

To get the upper bound, we proceed as follows

$$\begin{aligned}
[K:\mathbb{Q}]h(f) &= \sum_{v\in M_K} \max\{\log\|f_1\cdots f_m\|_v, 0\}\\
&= \sum_{v\in M_K^0} \max\left\{\sum_{j=1}^m \log\|f_j\|_v, 0\right\} + \sum_{v|\infty}[K_v:\mathbb{R}]\log^+|f_1\cdots f_m|_v \quad \text{by Gauß's Lemma (Lemma 1.3.5)}\\
&\le \sum_{v\in M_K^0}\sum_{j=1}^m \log^+\|f_j\|_v + \sum_{v|\infty}\max\left\{[K_v:\mathbb{R}]\log|f_1\cdots f_m|_v, 0\right\}\\
&\le \sum_{v\in M_K^0}\sum_{j=1}^m \log^+\|f_j\|_v + \sum_{v|\infty}\max\left\{[K_v:\mathbb{R}]\left(\sum_{j=1}^m\log|f_j|_v + d\log 2\right), 0\right\} \quad \text{by Gelfond's Lemma (Lemma 1.3.6)}\\
&= \sum_{v\in M_K^0}\sum_{j=1}^m \log^+\|f_j\|_v + \sum_{v|\infty}\max\left\{\sum_{j=1}^m\log\|f_j\|_v + [K_v:\mathbb{R}]d\log 2, 0\right\}\\
&\le \sum_{v\in M_K^0}\sum_{j=1}^m \log^+\|f_j\|_v + \sum_{v|\infty}\max\left\{\sum_{j=1}^m\log\|f_j\|_v, 0\right\} + \sum_{v|\infty}[K_v:\mathbb{R}]d\log 2\\
&\le \sum_{v\in M_K^0}\sum_{j=1}^m \log^+\|f_j\|_v + \sum_{v|\infty}\sum_{j=1}^m\log^+\|f_j\|_v + \sum_{v|\infty}[K_v:\mathbb{R}]d\log 2\\
&= [K:\mathbb{Q}]\sum_{j=1}^m h(f_j) + [K:\mathbb{Q}]d\log 2.
\end{aligned}$$

The "Moreover" part holds true because of the "Moreover" part of Gelfond's Lemma.

To get the lower bound, we have

$$\begin{aligned}
[K:\mathbb{Q}]h(f) &\ge \sum_{v\in M_K}\log\|f\|_v\\
&= \sum_{v\in M_K}\log\|f_1\cdots f_m\|_v\\
&= \sum_{v\in M_K^0}\sum_{j=1}^m\log\|f_j\|_v + \sum_{v|\infty}[K_v:\mathbb{R}]\log|f_1\cdots f_m|_v \quad \text{by Gauß's Lemma (Lemma 1.3.5)}\\
&\ge \sum_{j=1}^m\sum_{v\in M_K^0}\log\|f_j\|_v + \sum_{v|\infty}[K_v:\mathbb{R}](\sum_{j=1}^m\log|f_j|_v - d\log 2) \quad \text{by Gelfond's Lemma (Lemma 1.3.6)}\\
&= \sum_{j=1}^m\sum_{v\in M_K}\log\|f_j\|_v + \sum_{v|\infty}[K_v:\mathbb{R}]d\log 2\\
&= [K:\mathbb{Q}]\sum_{j=1}^m h(f_j) + [K:\mathbb{Q}]d\log 2.
\end{aligned}$$

We are done. □

So in the rest, we aim to prove Gelfond's Lemma (Lemma 1.3.6).

Definition 1.3.7. *The **Mahler measure** of f is defined to be*

$$M(f) := \exp\left(\int_{\mathbb{T}^n}\log|f(e^{i\theta_1},\dots,e^{i\theta_n})|\mathrm{d}\mu_1\cdots\mathrm{d}\mu_n\right),$$

where $\mathbb{T}$ *is the unit circle* $\{e^{i\theta} : 0 \le \theta < 2\pi\}$ *in* $\mathbb{R}$ *equipped with the standard measure* $\mathrm{d}\mu = (1/2\pi)\mathrm{d}\theta$.

The following *multiplicative* property of the Mahler measure is easy to check:

$$M(fg) = M(f)M(g). \tag{1.3.5}$$

Definition 1.3.8. *The* $\mathbf{L^2}$***-norm*** *of* f *is defined to be*

$$\ell_2(f) := \left(\int_{\mathbb{T}^n} |f(e^{i\theta_1}, \dots, e^{i\theta_n})|^2 \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_n\right)^{1/2} = \left(\sum_{\mathbf{j}} |a_{\mathbf{j}}|^2\right)^{1/2}.$$

In fact, we have given two equivalent definitions of the L^2-norm above. They coincide by Parseval's identity.

One-variable case We start by studying the one-variable case. The following lemma is an elementary tool to study the Mahler measure.

Lemma 1.3.9 (Jensen's Lemma)**.** *Let* $f(t) = a_d t^d + \cdots + a_0 \in \mathbb{C}[t]$. *Write* $\alpha_1, \dots, \alpha_d \in \mathbb{C}$ *for the roots of* f, *i.e.* $f(t) = a_d(t-\alpha_1)\cdots(t-\alpha_d)$. *Then we have*

$$\log M(f) = \log|a_d| + \sum_{j=1}^{d} \log^+|\alpha_j|,$$

with $\log^+(x) := \max\{\log x, 0\}$.

Proof. We only give a sketch here.

Because Mahler measure is multiplicative, it suffices to prove $\log M(t-\alpha) = \log^+|\alpha|$ for each $\alpha \in \mathbb{C}$.

If $|\alpha| > 1$, then the function $\log|t-\alpha|$ is harmonic in the unit disk, and hence its mean value on the unit circle is its value at the center which is $\log|\alpha| = \log^+|\alpha|$. If $|\alpha| < 1$, then the function $\log|1-\alpha\bar{t}|$ is harmonic in the unit disk and coincides with $\log|t-\alpha|$ on the unit circle, while its value at the center is $0 = \log^+|\alpha|$. Finally, the case $|\alpha| = 1$ is obtained by continuity. □

The following lemma uses the Mahler measure $M(f)$ to bound $\ell_\infty(f)$.

Lemma 1.3.10. *Let* $f(t) = a_d t^d + \cdots + a_0 \in \mathbb{C}[t]$. *Then we have*

$$\binom{d}{\lfloor d/2 \rfloor}^{-1} \ell_\infty(f) \le M(f) \le \ell_2(f) \le (d+1)^{1/2}\ell_\infty(f).$$

Proof. The last inequality is easy to see because $\ell_2(f) = (\sum_{j=0}^d |a_j|^2)^{1/2} \le (d+1)^{1/2}\max_j\{|a_j|\} = (d+1)^{1/2}\ell_\infty(f)$.

To prove the first inequality, write $f(t) = a_d(t-\alpha_1)\cdots(t-\alpha_d)$. Then for each $r \in \{0, \dots, d\}$ we have

$$|a_{d-r}| = |a_d|\left|\sum_{j_1<\cdots<j_r} \alpha_{j_1}\cdots\alpha_{j_r}\right| \le \binom{d}{r}|a_d|\prod_{j=1}^{d}\max\{1, |\alpha_j|\}.$$

Thus Jensen's Lemma above yields

$$|a_{d-r}| \le \binom{d}{r} M(f) \le \binom{d}{\lfloor d/2 \rfloor} M(f)$$

for each $r \in \{0, \dots, d\}$. So we have $\ell_\infty(f) \le \binom{d}{\lfloor d/2 \rfloor} M(f)$ and this is the first inequality.

To prove the inequality in the middle, we use *Jensen's inequality* which applies to convex functions. It says: If Ω is a space with a measure $\mathrm{d}\mu$ such that $\mathrm{d}\mu(\Omega) = \int_\Omega \mathrm{d}\mu = 1$, if g is a real-valued μ-integrable function on Ω and φ is a convex function on $\mathbb{R}$, then we have

$$\varphi\left(\int_\Omega g \mathrm{d}\mu\right) \le \int_\Omega (\varphi \circ g) \mathrm{d}\mu. \tag{1.3.6}$$

Applying this to $\Omega = \mathbb{T}$, $\mathrm{d}\mu$ as in the definition of Mahler measure and L^2-norm, $\varphi = \exp$ and $g(t) = 2\log|f(e^{i\theta})|$, we obtain

$$M(f)^2 \le \int_{\mathbb{T}} |f(e^{i\theta})|^2 \mathrm{d}\mu = \ell_2(f)^2.$$

Hence we are done for the middle inequality. □

Multi-variable case Here is the multi-variable version of the bound of $\ell_\infty(f)$ by $M(f)$.

Lemma 1.3.11. *Let $f(t_1, \dots, t_n) \in \mathbb{C}[t_1, \dots, t_n]$ with partial degrees $d_1, \dots, d_n$. Then*

$$\prod_{j=1}^n (d_j + 1)^{-1/2} M(f) \le \ell_\infty(f) \le \prod_{j=1}^n \binom{d_j}{\lfloor d_j/2 \rfloor} M(f).$$

Proof. The desired inequality is equivalent to

$$\prod_{j=1}^n \binom{d_j}{\lfloor d_j/2 \rfloor}^{-1} \ell_\infty(f) \le M(f) \le \prod_{j=1}^n (d_j + 1)^{1/2} \ell_\infty(f).$$

The proof for the second inequality follows the same line as in the one-variable case; one uses the L^2-norm as an intermediate. More precisely, one uses *Jensen's inequality* (1.3.6) to prove $M(f) \le \ell_2(f)$, and then applies the easy bound $\ell_2(f) = (\sum_{1 \le j \le d,\ 0 \le i_j \le d_j} |a_{i_1,\dots,i_d}|^2)^{1/2} \le \prod_{j=1}^n (d_j + 1)^{1/2} \ell_\infty(f)$.

Now we prove the first inequality $\prod_{j=1}^n \binom{d_j}{\lfloor d_j/2 \rfloor}^{-1} \ell_\infty(f) \le M(f)$ by induction on n. The base step $n = 1$ is proved in Lemma 1.3.10.

Assume the result is proved for $1, \dots, n-1$. We can write uniquely

$$f(t_1, \dots, t_n) = \sum_{j=0}^{d_n} f_j(t_1, \dots, t_{n-1}) t_n^j$$

for certain polynomials $f_j \in \mathbb{C}[t_1, \dots, t_{n-1}]$. Then $\ell_\infty\left(f(e^{i\theta_1}, \dots, e^{i\theta_{n-1}}, t)\right) = \max_j |f_j(e^{i\theta_1}, \dots, e^{i\theta_{n-1}})|$.

Fixing $\theta_1, \dots, \theta_{n-1}$, we have

$$\log M\left(f(e^{i\theta_1}, \dots, e^{i\theta_{n-1}}, t)\right) = \int_{\mathbb{T}} \log|f(e^{i\theta_1}, \dots, e^{i\theta_n})| \mathrm{d}\mu_n,$$

and thus

$$\begin{aligned} \log M(f) &= \int_{\mathbb{T}^n} \log|f(e^{i\theta_1}, \dots, e^{i\theta_n})| \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_n \\ &= \int_{\mathbb{T}^{n-1}} \log M\left(f(e^{i\theta_1}, \dots, e^{i\theta_{n-1}}, t)\right) \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{n-1}. \end{aligned}$$

Fixing $\theta_1, \dots, \theta_{n-1}$, we apply the first inequality in Lemma 1.3.10 to the one-variable polynomial $f(e^{i\theta_1}, \dots, e^{i\theta_{n-1}}, t)$. We then get

$$M\left(f(e^{i\theta_1}, \dots, e^{i\theta_{n-1}}, t)\right) \geq \binom{d_n}{\lfloor d_n/2 \rfloor}^{-1} \max_j |f_j(e^{i\theta_1}, \dots, e^{i\theta_{n-1}})|.$$

Thus we have

$$\begin{aligned}
\log M(f) &\geq \int_{\mathbb{T}^{n-1}} \log \max_j |f_j(e^{i\theta_1}, \dots, e^{i\theta_{n-1}})| \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{n-1} - \log \binom{d_n}{\lfloor d_n/2 \rfloor} \\
&\geq \max_j \int_{\mathbb{T}^{n-1}} \log |f_j(e^{i\theta_1}, \dots, e^{i\theta_{n-1}})| \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{n-1} - \log \binom{d_n}{\lfloor d_n/2 \rfloor} \\
&= \max_j \log M(f_j) - \log \binom{d_n}{\lfloor d_n/2 \rfloor} \\
&\geq \max_j \log \ell_\infty(f_j) - \sum_{j=1}^{n} \log \binom{d_j}{\lfloor d_j/2 \rfloor} \quad \text{by induction hypothesis} \\
&= \log \ell_\infty(f) - \sum_{j=1}^{n} \log \binom{d_j}{\lfloor d_j/2 \rfloor}.
\end{aligned}$$

This is what we desire. We are done. □

Now we are ready to prove *Gelfond's Lemma*.

Proof of Lemma 1.3.6. Recall the set-up. We have $f_1, \dots, f_m \in \mathbb{C}[t_1, \dots, t_n]$ and $f := f_1 \cdots f_m$. Let d be the sum of the partial degrees of f. We wish to prove

$$2^{-d} \prod_{j=1}^{m} \ell_\infty(f_j) \leq \ell_\infty(f) \leq 2^d \prod_{j=1}^{m} \ell_\infty(f_j).$$

Write $d_1^{(j)}, \dots, d_n^{(j)}$ for the partial degrees of f_j.

We start with the lower bound for $\ell_\infty(f)$. The proof uses the relation between $M(f)$ and $\ell_\infty(f)$ established in Lemma 1.3.11. Recall that $M(f) = M(f_1) \cdots M(f_m)$. We have

$$\begin{aligned}
\prod_{j=1}^{m} \ell_\infty(f_j) &\leq \prod_{j=1}^{m} \left(\prod_{k=1}^{n} \binom{d_k^{(j)}}{\lfloor d_k^{(j)}/2 \rfloor} M(f_j) \right) \quad \text{by the second inequality in Lemma 1.3.11} \\
&= \prod_{j=1}^{m} \prod_{k=1}^{n} \binom{d_k^{(j)}}{\lfloor d_k^{(j)}/2 \rfloor} M(f) \\
&\leq \left(\prod_{j=1}^{m} \prod_{k=1}^{n} \binom{d_k^{(j)}}{\lfloor d_k^{(j)}/2 \rfloor} \right) \left(\prod_{k=1}^{n} \left(1 + \sum_{j=1}^{m} d_k^{(j)} \right)^{1/2} \right) \ell_\infty(f) \quad \text{by the first inequality in Lemma 1.3.11.}
\end{aligned}$$

Then the upper bound is obtained from the following fact: Let $a \leq A$, $b \leq B$ and d be non-negative integers. Then $\binom{A}{a}\binom{B}{b} \leq \binom{A+B}{a+b}$ and $\binom{d}{\lfloor d/2 \rfloor}(d+1)^{1/2} \leq 2^d$.[5]

[5]The first follows from $(1+t)^A(1+t)^B = (1+t)^{A+B}$, and the second follows from Stirling's formula.

Next we prove the upper bound for $\ell_\infty(f)$. For this, we will establish

$$\ell_\infty(f) \le C \prod_{j=1}^{m} \ell_\infty(f_j)$$

with

$$C = \prod_{j=1}^{m-1} \prod_{k=1}^{n} \left(1 + d_k^{(j)}\right) \le 2^d. \tag{1.3.7}$$

Let us explain how this C is chosen. First notice that *only the degrees of the first $m-1$ polynomials count*. This observation is in many applications important. It also gives the "Moreover" part of Gelfond's Lemma.

Write $f_j = \sum_{\mathbf{k}} a_{\mathbf{k}}^{(j)} \mathbf{t}^{\mathbf{k}} = \sum_{0 \le k_1 \le d_1^{(j)}, \dots, 0 \le k_n \le d_n^{(j)}} a_{k_1,\dots,k_n}^{(j)} t_1^{k_1} \cdots t_n^{k_n}$. Then

$$\begin{aligned} f = \prod_{j=1}^{m} f_j &= \prod_{j=1}^{m} \left(\sum_{\mathbf{k}} a_{\mathbf{k}}^{(j)} \mathbf{t}^{\mathbf{k}} \right) \\ &= \sum_{\mathbf{e}} \left(\sum_{\mathbf{k}^{(1)} + \cdots + \mathbf{k}^{(m)} = \mathbf{e}} a_{\mathbf{k}^{(1)}}^{(1)} \cdots a_{\mathbf{k}^{(m)}}^{(m)} \right) \mathbf{t}^{\mathbf{e}}. \end{aligned}$$

Here $\mathbf{e} = (e_1, \dots, e_n)$ is a multi-index with n components, and each $\mathbf{k}^{(j)} = (k_1^{(j)}, \dots, k_n^{(j)})$ is also a multi-index with n components. Moreover, we have $0 \le k_1^{(j)} \le d_1^{(j)}, \dots, 0 \le k_n^{(j)} \le d_n^{(j)}$.

Now we are reduced to the following claim: For each fixed $\mathbf{e}$, we need to prove that the number of monomials in $\sum_{\mathbf{k}^{(1)} + \cdots + \mathbf{k}^{(m)}} a_{\mathbf{k}^{(1)}}^{(1)} \cdots a_{\mathbf{k}^{(m)}}^{(m)}$ is at most C. Notice that under this assumption, if $\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(m-1)}$ are all fixed, then $\mathbf{k}^{(m)}$ is also fixed. Hence we can conclude because C is the naive upper bound for the number of choices of the tuple $(\mathbf{k}^{(1)}, \dots, \mathbf{k}^{(m-1)})$ satisfying that $0 \le k_1^{(j)} \le d_1^{(j)}, \dots, 0 \le k_n^{(j)} \le d_n^{(j)}$. □

1.3.3 Some other operations with polynomials

The contents below are covered in the Exercise class.

We have seen how to bound the height of the product of polynomials. Now we turn to other operations.

The first is the *sum* of polynomials. For this, Proposition 1.2.8 implies the following bound rather easily.

Proposition 1.3.12. *Let $f_1, \dots, f_r \in \overline{\mathbb{Q}}[t_1, \dots, t_n]$. Then we have*

$$h(f_1 + \ldots + f_r) \le \sum_{j=1}^{r} h(f_j) + \log r.$$

In what follows in this subsection, let $f(\mathbf{t}) = \sum_{\mathbf{j}} \mathbf{t}^{\mathbf{j}} = \sum_{j_1,\dots,j_n} a_{j_1 \dots j_n} t_1^{j_1} \cdots t_n^{j_n}$.

Next, we turn to the formal partial derivatives $\partial f / \partial t_k := \sum_{j_1,\dots,j_n,\ j_k \ge 1} j_k a_{j_1 \dots j_n} t_1^{j_1} \cdots t_{k-1}^{j_{k-1}} t_k^{j_k - 1} t_{k+1}^{j_{k+1}} \cdots t_n^{j_n}$.

Proposition 1.3.13. *Let $d_{\max}$ be the maximum of the partial degrees of f. Then*

$$h\left(\frac{\partial f}{\partial t_k}\right) \le h(f) + \log d_{\max}.$$

Proof. Let K be a number field such that all the coefficients of f are in K.

Each $j_k \neq 0$ appearing in the monomials of f satisfies $1 \leq j_k \leq d_{\max}$. If v is non-archimedean, then $\|j_k\|_v \leq 1$. If v is archimedean, then $\|j_k\|_v \leq \|d_{\max}\|_v$. In summary, $\|j_k\|_v \leq \max\{1, \|d_{\max}\|_v\}$ for each $v \in M_K$.

Notice that each coefficient of $\partial f/\partial t_k$ is $j_k a_{j_1 \dots j_n}$. Thus

$$\|\partial f/\partial t_k\|_v \leq \max\{1, \|d_{\max}\|_v\}\|f\|_v,$$

and hence $\max\{1, \|\partial f/\partial t_k\|_v\} \leq \max\{1, \|d_{\max}\|_v\} \max\{1, \|f\|_v\}$. So

$$\begin{aligned}
[K:\mathbb{Q}]h(\partial f/\partial t_k) &= \sum_{v \in M_K} \log^+ \|\partial f/\partial t_k\|_v \\
&\leq \sum_{v \in M_K} \log^+ \|d_{\max}\|_v + \sum_{v \in M_K} \log^+ \|f\|_v \\
&= [K:\mathbb{Q}]h(d_{\max}) + [K:\mathbb{Q}]h(f).
\end{aligned}$$

Hence $h(\partial f/\partial t_k) \leq \log d_{\max} + h(f)$ because $h(d_{\max}) = \log d_{\max}$ as $d_{\max}$ is a positive integer. □

1.3.4 Mahler measure and algebraic number

Let us see another application of Jensen's Lemma (Lemma 1.3.9)[6], which establishes the relation between the Mahler measure and the height of an algebraic number.

Proposition 1.3.14. *Let $\alpha \in \overline{\mathbb{Q}}$ and let f be the minimal polynomial of α over $\mathbb{Z}$. Then we have*

$$\log M(f) = \deg(\alpha) h(\alpha). \tag{1.3.8}$$

In particular, we have

$$\log |N_{\mathbb{Q}(\alpha)/\mathbb{Q}}(\alpha)| \leq \deg(\alpha) h(\alpha). \tag{1.3.9}$$

Proof. Set $d = \deg(\alpha)$ and write $f(t) = a_d t^d + \cdots + a_0 \in \mathbb{Z}[t]$. Write $\alpha_1 = \alpha, \dots, \alpha_d$ the Galois conjugates of α. Then $f(t) = a_d(t-\alpha_1)\cdots(t-\alpha_d)$. Let $K \subseteq \overline{\mathbb{Q}}$ be the Galois closure of $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$, *i.e.* K is the smallest Galois extension over $\mathbb{Q}$ which contains α and all its Galois conjugate. Write $G = \mathrm{Gal}(K/\mathbb{Q})$. Then $\{\sigma(\alpha)\}_{\sigma \in G}$ contains every conjugate of α exactly $[K:\mathbb{Q}]/d$ times.

For each (non-archimedean) $v \in M_K^0$, Gauß's Lemma (Lemma 1.3.5) yields

$$\max\{\|a_d\|_v, \dots, \|a_0\|_v\} = \|f\|_v = \|a_d\|_v \prod_{i=1}^{d} \max\{1, \|\alpha_i\|_v\}.$$

Notice that the left hand side equals 1 because each $a_i \in \mathbb{Z}$ and $\gcd(a_d, \dots, a_0) = 1$. Thus

$$\log \|a_d\|_v + \sum_{i=1}^{d} \log^+ \|\alpha_i\|_v = 0 \tag{1.3.10}$$

for each $v \in M_K^0$.

[6] $\log M(f) = \log |a_d| + \sum_{j=1}^{d} \log^+ |\alpha_j|$ for $f(t) = a_d(t-\alpha_1)\cdots(t-\alpha_d)$.

Now we have

$$\begin{aligned}
[K:\mathbb{Q}]h(\alpha) &= \frac{[K:\mathbb{Q}]}{d}\sum_{i=1}^{d} h(\alpha_i) \quad \text{by Lemma 1.2.4} \\
&= \frac{1}{d}\sum_{i=1}^{d}\sum_{v\in M_K} \log^+ \|\alpha_i\|_v \\
&= \frac{1}{d}\left(\sum_{v|\infty}\sum_{i=1}^{d}\log^+ \|\alpha_i\|_v - \sum_{v\in M_K^0}\log\|a_d\|_v\right) \quad \text{by (1.3.10)} \\
&= \frac{1}{d}\sum_{v|\infty}\left(\log\|a_d\|_v + \sum_{i=1}^{d}\log^+\|\alpha_i\|_v\right) \quad \text{by Product Formula applied to } a_d.
\end{aligned}$$

If $K_v = \mathbb{R}$, then $\|\cdot\|_v = |\cdot|$. If $K_v = \mathbb{C}$, then $\|\cdot\|_v = |\cdot|^2$. Recall, from Algebraic Number Theory, the basic fact that $\#\{v : K_v = \mathbb{R}\} + 2\#\{v : K_v = \mathbb{C}\} = [K:\mathbb{Q}]$. Hence we can apply Jensen's Lemma (Lemma 1.3.9) to each term on the right hand side and obtain

$$[K:\mathbb{Q}]h(\alpha) = \frac{[K:\mathbb{Q}]}{d}\log M(f).$$

This yields (1.3.8).

To prove the "In particular" part, recall from Algebraic Number Theory that $N_{\mathbb{Q}(\alpha)/\mathbb{Q}}(\alpha) = \prod_{i=1}^{d}\alpha_i$. Thus $\log|N_{\mathbb{Q}(\alpha)/\mathbb{Q}}(\alpha)| = \sum_{i=1}^{d}\log|\alpha_i|$, which then $\le \sum_{i=1}^{d}\log^+|\alpha_i|$ and hence $\le \log M(f)$ by Jensen's Lemma. □

Remark 1.3.15. *Let us have a quick look at the* ***Lehmer Conjecture****. If $\alpha \neq 0$ is an algebraic number with minimal polynomial f, the Mahler measure of α is defined to be $M(\alpha) := M(f)$. Then* (1.3.8) *yields $M(\alpha) = H(\alpha)^{\deg(\alpha)}$. Lehmer's conjecture predicts that there exists a contant c such that $M(\alpha) \ge c > 1$ for all $\alpha \in \overline{\mathbb{Q}}^*$ not a root of unity. Alternatively, $h(\alpha) \ge c/\deg(\alpha)$ for some absolute constant c. This conjecture is open. Currently, we have Dobrowolski's theorem which claims ($d := \deg(\alpha)$)*

$$M(\alpha) \ge 1 + c\left(\frac{\log\log d}{\log d}\right)^3.$$

Chapter 2

Siegel Lemma

2.1 Basic version

We start with the very basic version of Siegel's Lemma.

Lemma 2.1.1. *Let $a_{ij} \in \mathbb{Z}$ with $i = 1, \dots, M$ and $j = 1, \dots, N$. Assume that a_{ij} are not all 0 and $|a_{ij}| \le B$ for all i and j.*

If $N > M$, then the homogeneous linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N &= 0 \end{aligned}$$

has a non-zero solution $(x_1, \dots, x_N) \in \mathbb{Z}^N$ with

$$\max_j |x_j| \le \lfloor (NB)^{\frac{M}{N-M}} \rfloor.$$

In practice, it is more convenient to denote by $A = (a_{ij})_{1 \le i \le M,\ 1 \le j \le N}$ which is a non-zero $M \times N$-matrix with entries in $\mathbb{Z}$. The upshot of this lemma is that the linear system $A\mathbf{x} = \mathbf{0}$ has a *small* non-zero solution provided that $N > M$. Here *small* means that the height of this non-zero solution is bounded in terms of N, M and $h(A)$.[1] It should be understood that *M is the number of equations and $N - M$ is the dimension of the space of solutions.*

Proof. We may and do assume that no row of A is identically 0. Thus $M \ge 1$. For a positive integer k, consider the set

$$T := \{\mathbf{x} \in \mathbb{Z}^N : 0 \le x_j \le k,\ j = 1, \dots, N\}.$$

Then $\#T = (k+1)^N$.

Next, for each $i \in \{1, \dots, M\}$, denote by S_i^+ the sum of the positive entries in the i-th row of A, and by S_i^- the sum of the negative entries. Then

$$\text{For } \mathbf{x} \in T \text{ and } \mathbf{y} := A\mathbf{x}, \text{ we have } kS_i^- \le y_i \le kS_i^+ \quad \text{for each } i. \tag{2.1.1}$$

Next, set

$$T' := \{\mathbf{y} \in \mathbb{Z}^N : kS_i^- \le y_i \le kS_i^+ \text{ for each } i\}.$$

[1] Here $h(A)$ is defined to be the height of $[a_{ij}]_{i,j}$ viewed as a point in $\mathbb{P}^{MN-1}(\overline{\mathbb{Q}})$.

Then for $B_i := \max_j |a_{ij}|$, we have $S_i^+ - S_i^- \leq NB_i$ and we can conclude that $\#T' \leq \prod_{i=1}^M (NkB_i + 1)$.

Now take $k := \lfloor \prod_{i=1}^M (NB_i)^{1/(N-M)} \rfloor$. Then $NkB_i + 1 < NB_i(k+1)$ because $N \geq M > 1$, and hence

$$\prod_{i=1}^M (NkB_i + 1) < \prod_{i=1}^M NB_i(k+1) = (k+1)^M \prod_{i=1}^M NB_i.$$

On the other hand, $\prod_{i=1}^M (NB_i)^{1/(N-M)} \leq k+1$. So

$$\prod_{i=1}^M (NkB_i + 1) < (k+1)^M (k+1)^{N-M} = (k+1)^N = \#T.$$

We have seen that $\#T'$ is bounded above by the left hand side. So $\#T' < \#T$. By the Pigeonhole Principle and (2.1.1), there exist two different points $\mathbf{x}', \mathbf{x}'' \in T$ such that $A\mathbf{x}' = A\mathbf{x}''$.

Now $\mathbf{x} := \mathbf{x}' - \mathbf{x}''$ is a non-zero solution of the linear system in question such that $\max_j |x_j| \leq k = \lfloor \prod_{i=1}^M (NB_i)^{1/(N-M)} \rfloor \leq \lfloor (NB)^{M/(N-M)} \rfloor$. □

This basic version self-improves to a version for number fields.

Lemma 2.1.2. *Let $K \subseteq \mathbb{C}$ be a number field of degree d, and let $|\cdot|$ be the usual absolute value on $\mathbb{C}$. Let $M, N \in \mathbb{Z}$ with $0 < M < N$. Then there exist positive integers C_1 and C_2 such that the following property holds true: For any non-zero $M \times N$-matrix A with entries $a_{mn} \in \mathcal{O}_K$, there exists $\mathbf{x} \in \mathcal{O}_K^N \setminus \{\mathbf{0}\}$ with $A\mathbf{x} = \mathbf{0}$ and*

$$H(\mathbf{x}) \leq C_1 (C_2 NB)^{\frac{M}{N-M}},$$

where $B := \max_{\sigma,m,n} |\sigma(a_{mn})|$ with σ running over all the embeddings $K \hookrightarrow \mathbb{C}$.

The constants C_1 and C_2 depend only K (and hence d), M and N, but they are independent of the choice of the matrix A.[2] By the Fundamental Inequality (Proposition 1.2.10), B can be bounded by $H(A)$ with A viewed as a point $(a_{mn})_{m,n} \in \overline{\mathbb{Q}}^{MN}$.

Proof. Let $\omega_1, \ldots, \omega_d$ be a $\mathbb{Z}$-basis of $\mathcal{O}_K$. The entries of A may be written as

$$a_{mn} = \sum_{j=1}^d a_{mn}^{(j)} \omega_j, \quad a_{mn}^{(j)} \in \mathbb{Z}. \tag{2.1.2}$$

For each $\mathbf{x} = (x_1, \ldots, x_N) \in \mathcal{O}_K^N$, using $x_n = \sum_{k=1}^d x_n^{(k)} \omega_k$ we get

$$(A\mathbf{x})_m = \sum_{n=1}^N \sum_{j,k=1}^d a_{mn}^{(j)} \omega_j \omega_k x_n^{(k)} = \sum_{l=1}^d \sum_{n=1}^N \sum_{j,k=1}^d a_{mn}^{(j)} b_{jk}^{(l)} x_n^{(k)} \omega_l,$$

where $\omega_j \omega_k = \sum_{l=1}^d b_{jk}^{(l)} \omega_l$. Set A' to be the $(Md) \times (Nd)$-matrix

$$A' := \left(\sum_{j=1}^d a_{mn}^{(j)} b_{jk}^{(l)} \right)$$

[2] In fact by the proof, one can see that C_2 depends only on K.

with rows indexed by (m,l) and columns indexed by (n,k). Write $\mathbf{y} \in \mathbb{Z}^{Nd}$ for the vector $(x_n^{(k)})$.

Apply the basic version of Siegel's Lemma, Lemma 2.1.1, to A'. Then we obtain a non-zero integer solution $\mathbf{y}$ with $A'\mathbf{y} = \mathbf{0}$ such that

$$H(\mathbf{y}) \leq \left(Nd^2 \max_{m,n,j} |a_{mn}^{(j)}| \max_{j,k,l} |b_{jk}^{(l)}| \right)^{\frac{M}{N-M}}.$$

As $x_n = \sum_{k=1}^{d} x_n^{(k)} \omega_k$ for each n, we then obtain a constant C_1 such that $H(\mathbf{x}) \leq C_1 H(\mathbf{y})$.

Next we wish to bound $\max_j |a_{mn}^{(j)}|$ in terms of $\max_{\sigma,m,n} |\sigma(a_{mn})|$. Let σ run over the $d = [K : \mathbb{Q}]$ different embeddings $K \hookrightarrow \mathbb{C}$. Apply each σ to (2.1.2). It is known from Algebraic Number Theory that the $d \times d$-matrix $(\sigma(\omega_j))_{\sigma,j}$ is invertible.[3] So we obtain a constant C_2' such that

$$\max_j |a_{mn}^{(j)}| \leq C_2' \max_{\sigma} |\sigma(a_{mn})|.$$

Thus we can conclude by taking $C_2 := C_2' d^2 \max_{j,k,l} |b_{jk}^{(l)}|$. □

Next, we also have the following *relative version* of Siegel's Lemma.

Lemma 2.1.3 (Relative version of Siegel's Lemma, basic version)**.** *Let K be a number field of degree d. Then there exists a positive number C such that the following property holds true For any $M, N \in \mathbb{Z}$ with $0 < dM < N$ and any non-zero $M \times N$-matrix A with entries $a_{mn} \in \mathcal{O}_K$, there exists $\mathbf{x} \in \mathbb{Z}^N \setminus \{0\}$ with $A\mathbf{x} = \mathbf{0}$ and*

$$H(\mathbf{x}) \leq \lfloor (CNB)^{\frac{dM}{N-dM}} \rfloor$$

where $B := \max_{\sigma,m,n} |\sigma(a_{mn})|$ with σ running over all the embeddings $K \hookrightarrow \mathbb{C}$.

Again, by the Fundamental Inequality (Proposition 1.2.10), B can be bounded by $H(A)$ with A viewed as a point $(a_{mn})_{m,n} \in \overline{\mathbb{Q}}^{MN}$. We emphasize that the constant C depends only on the field K.

Proof. Let $\omega_1, \ldots, \omega_d$ be a $\mathbb{Z}$-basis of $\mathcal{O}_K$. For the entries of $A = (a_{mn})$, we have

$$a_{mn} = \sum_{j=1}^{d} a_{mn}^{(j)} \omega_j \tag{2.1.3}$$

for uniquely determined $a_{mn}^{(j)} \in \mathbb{Z}$. Consider the $M \times N$-matrix $A^{(j)} = (a_{mn}^{(j)})$ for each $j \in \{1, \ldots, d\}$. Then for $\mathbf{x} \in \mathbb{Q}^N$, the equation $A\mathbf{x} = \mathbf{0}$ is equivalent to the system of equations $A^{(j)}\mathbf{x} = \mathbf{0}$ for all $j = 1, \ldots, d$. This new system has dM equations and N unknowns. Write A' for the $dM \times N$-matrix $\begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(d)} \end{pmatrix}$. Since $dM < N$, we can apply Lemma 2.1.1 to find a non-zero solution $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{Z}^N$ with

$$\max_i |x_i| \leq \lfloor (N \max_{m,n,j} |a_{mn}^{(j)}|)^{\frac{dM}{N-dM}} \rfloor.$$

It remains to compare $\max_{m,n,j} |a_{mn}^{(j)}|$ and $\max_{\sigma,m,n} |\sigma(a_{mn})|$. We use the same argument as for Lemma 2.1.2. Let σ run over the $d = [K : \mathbb{Q}]$ different embeddings $K \hookrightarrow \mathbb{C}$. Apply each σ to (2.1.3). It is known from Algebraic Number Theory that $\deg(\sigma(\omega_j))_{\sigma,j}^2 = \mathrm{Disc}(K/\mathbb{Q}) \neq 0$. So we obtain a constant C such that $\max_j |a_{mn}^{(j)}| \leq C \max_{\sigma} |\sigma(a_{mn})|$. Hence we are done. □

[3] $\deg(\sigma(\omega_j))_{\sigma,j}^2 = \mathrm{Disc}(K/\mathbb{Q}) \neq 0$.

2.2 Faltings's version of Siegel's Lemma

In his famous paper `Diophantine approximation on abelian varieties` (*Annals of Math.* **133**:549–576, 1991), Faltings proved a fancier Siegel's Lemma. It plays a fundamental role for his proof of the Mordell–Lang Conjecture. In this section, we discuss about this.

2.2.1 Background and statement

Recall the following basic version of Siegel's Lemma, Lemma 2.1.1.

Lemma 2.2.1. *Let $A = (a_{ij})$ be an $M \times N$-matrix with entries in $\mathbb{Z}$. Set $B = \max_{i,j} |a_{ij}|$. If $N > M$, then* $\mathrm{Ker}(A)$ *contains a non-zero vector $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{Z}^N$ such that*

$$\max_j |x_j| \leq (NB)^{\frac{M}{N-M}}.$$

Let us digest this lemma in the following way. The matrix A defines a linear map $\alpha\colon \mathbb{R}^N \to \mathbb{R}^M$ such that $\alpha(\mathbb{Z}^N) \subseteq \mathbb{Z}^M$, *i.e* α maps the lattice $\mathbb{Z}^N$ into the lattice $\mathbb{Z}^M$. If $N > M$, then we are able to find a non-trivial lattice point *of small norm* in $\mathrm{Ker}(\alpha)$. As we said before, $N - M$ should be understood to be $\dim \mathrm{Ker}(A)$ (although in the current formulation they may not be the same).

Faltings's fancier version looks not for only one, but for an *arbitrary number of linearly independent lattice points* in $\mathrm{Ker}(\alpha)$. To say that these lattice points are *of small norm*, we use the *successive minima*. Moreover, it is more natural to work with arbitrary normed real vector spaces.

Let $(V, \|\cdot\|)$ be a finite dimensional normed real vector space, and let Λ be a lattice (a discrete subgroup of V which spans V). Denote by $B(V)$ the unit ball $\{x \in V : \|x\| \leq 1\}$ in V.

Definition 2.2.2. *The n-**th successive minimum** of $(V, \|\cdot\|, \Lambda)$ is*

$$\begin{aligned}\lambda_n(V, \|\cdot\|, \Lambda) :&= \inf\{t > 0 : \ \Lambda \textit{ contains } n \textit{ linearly independent vectors of norm} \leq t\} \\ &= \inf\{t > 0 : tB(V) \textit{ contains } n \textit{ linearly-independent vectors of } \Lambda\}.\end{aligned}$$

Next for two normed real vector spaces $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$, the *norm* of a linear map $\alpha\colon V \to W$ is defined to be

$$\|\alpha\| := \sup\left\{\frac{\|\alpha(x)\|_W}{\|x\|_V} : x \neq 0\right\}. \tag{2.2.1}$$

We are ready to state Faltings's version of Siegel's Lemma.

Theorem 2.2.3. *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be two finite dimensional normed real vector spaces, let Λ_V be a lattice in V and Λ_W be a lattice in W.*

Let $\alpha\colon V \to W$ be a linear map with $\alpha(\Lambda_V) \subseteq \Lambda_W$. Assume furthermore that there exists a real number $C \geq 2$ such that

(i) $\|\alpha\| \leq C$,

(ii) Λ_V *is generated by elements of norm* $\leq C$,

(iii) every non-zero element of Λ_V and of Λ_W has norm $\geq C^{-1}$.

Then for $U := \mathrm{Ker}(\alpha)$ *with the induced norm* $\|\cdot\|_U$ *(the restriction of* $\|\cdot\|_V$ *on* U*) and the lattice* $\Lambda_U := \Lambda_V \cap U$*, we have*

$$\lambda_{n+1}(U, \|\cdot\|_U, \Lambda_U) \leq \Big(C^{3\dim V} \cdot (\dim V)!\Big)^{1/(\dim U - n)}$$

for each $0 \leq n \leq \dim U - 1$.

Notice that the hypotheses (i)–(iii) can always be achieved by enlarging C.

The basic version of Siegel's Lemma (Lemma 2.2.1), up to changing the constant, follows from Theorem 2.2.3 with $n = 0$.

2.2.2 Proof of Theorem 2.2.3

The proof of Theorem 2.2.3 uses Minkowski's Second Theorem.

Let $(V, \|\cdot\|_V, \lambda_V)$ be a finite dimensional normed real vector space with a lattice. Set $d_V := \dim V$. For simplicity, denote by

$$V/\Lambda_V := \{v \in V : v = \sum_{j=1}^{d_V} \lambda_j v_j,\ 0 \leq \lambda_j < 1\}$$

where $\{v_1, \ldots, v_{d_V}\}$ is a basis of Λ_V. Notice that V/Λ_V depends on the choice of the basis.

We can endow V with a Lebesgue measure μ_V as follows. Fix an isomorphism $\psi\colon V \simeq \mathbb{R}^{d_V}$ and use μ to denote the standard Lebesgue measure on $\mathbb{R}^{d_V}$. Then set for any Lebesgue measurable $A \subseteq \mathbb{R}^{d_V}$

$$\mu_V(\psi^{-1}(A)) = \mu(A). \tag{2.2.2}$$

Up to a constant, there is only one Lebesgue measure on V. Thus the quantity

$$\mathrm{Vol}(V) = \mathrm{Vol}(V, \|\cdot\|_V, \Lambda_V) := \frac{\mu_V(B(V))}{\mu_V(V/\Lambda_V)} \tag{2.2.3}$$

does not depend on the choice of μ_V; it clearly does not depend on the choice of the basis of Λ_V in the definition of V/Λ_V.

Theorem 2.2.4 (Minkowski's Second Theorem)**.** *With the notation above, we have*

$$\frac{2^{d_V}}{d_V!} \leq \prod_{n=1}^{d_V} \lambda_n(V, \|\cdot\|_V, \Lambda_V) \cdot \mathrm{Vol}(V) \leq 2^{d_V}.$$

Here we used the fact that the unit ball $B(V)$ is convex and symmetric (*i.e.* $B(V) = -B(V)$).

To apply Minkowski's Second Theorem to prove Theorem 2.2.3, we need one last preparation on the *quotient norm*. More precisely, on V/U, we consider the norm

$$\|\overline{v}\|_{V/U} := \inf\{\|v + u\|_V : u \in U\}$$

for each $v \in V$. Having this norm, we can define the unit ball $B(V/U)$. Moreover, $\alpha(\Lambda_V)$ is a lattice in $\alpha(V)$, which can then be viewed as a lattice in V/U by the natural isomorphism $V/U \simeq \alpha(V)$. So we can define $\mathrm{Vol}(V/U) := \mathrm{Vol}(V/U, \|\cdot\|_{V/U}, \alpha(\Lambda_V))$. Recall the notation from Theorem 2.2.3; we naturally have the quantity $\mathrm{Vol}(U) := \mathrm{Vol}(U, \|\cdot\|_U, \Lambda_U)$.

Lemma 2.2.5. $\mathrm{Vol}(V) \leq 2^{\dim U}\mathrm{Vol}(U)\mathrm{Vol}(V/U)$.

Proof of Theorem 2.2.3 assuming Lemma 2.2.5. We will identify $V/U \simeq \alpha(V)$ in the proof. Take $w \in \alpha(\Lambda_V) \setminus \{0\}$. Write $w = \alpha(v)$ for some $v \in \Lambda_V$. Then

$$\|w\|_{V/U} = \inf_{u \in U} \|v + u\|_V \geq \frac{\|\alpha(v)\|_W}{\|\alpha\|} \geq C^{-2};$$

here the last inequality follows from hypotheses (i) and (iii). In particular, this implies that $\lambda_1(V/U, \|\cdot\|_{V/U}, \alpha(\Lambda_V)) \geq C^{-2}$.

Write $d_V := \dim V$ and $d_U := \dim U$. Minkowski's Second Theorem (applied to V/U) yields $\lambda_1(V/U, \|\cdot\|_{V/U}, \alpha(\Lambda_V))^{\dim V/U} \cdot \mathrm{Vol}(V/U) \leq 2^{\dim V/U}$. Thus from the paragraph above, we get $\mathrm{Vol}(V/U) \leq (2C^2)^{d_V - d_U}$.

Next, by hypothesis (ii), we have $\lambda_{d_V}(V, \|\cdot\|_V, \Lambda_V) \leq C$. Thus Minkowski's Second Theorem (applied to V) yields $\mathrm{Vol}(V) \geq 2^{d_V} C^{-d_V}/d_V!$.

Apply Lemma 2.2.5 and the volume estimates above. Then we get

$$\mathrm{Vol}(U)^{-1} \leq C^{3d_v - 2d_U} \cdot d_V!. \tag{2.2.4}$$

We apply another time Minkowski's Second Theorem (to U). For each $0 \leq n \leq d_U - 1$, we then get $\lambda_1(U, \|\cdot\|_U, \Lambda_U)^n \cdot \lambda_{n+1}(U, \|\cdot\|_U, \Lambda_U)^{d_U - n} \cdot \mathrm{Vol}(U) \leq 2^{d_U}$. But $\lambda_1(U, \|\cdot\|_U, \Lambda_U) \geq C^{-1}$ by hypothesis (iii). So we obtain

$$\begin{aligned}
\lambda_{n+1}(U, \|\cdot\|_U, \Lambda_U) &\leq \left(2^{d_U} \mathrm{Vol}(U)^{-1} C^n\right)^{1/(d_U - n)} \\
&\leq \left(2^{d_U} C^{n + 3d_v - 2d_U} \cdot d_V!\right)^{1/(d_U - n)} \qquad \text{by (2.2.4)} \\
&\leq \left(C^{3d_V} \cdot d_V!\right)^{1/(d_U - n)}.
\end{aligned}$$

Hence we are done. □

Proof of Lemma 2.2.5. Write $d_U := \dim U$ and $d_V := \dim V$.

Let μ_V and μ_U be the Lebesgue measures on V and U, respectively. On V/U we have a unique Lebesgue measure $\mu_{V/U}$ determined as follows: For any μ_V-measurable subset $E \subseteq V$, we have

$$\mu_V(E) = \int_{V/U} f_E(\overline{v}) \mathrm{d}\mu_{V/U}(\overline{v})$$

where $f_E(\overline{v}) := \mu_U(\{u \in U : u + v \in E\})$; here $f_E(\overline{v})$ is independent of the representative v because μ_U is translation invariant.

We compute $f_{B(V)}(\overline{v})$ for $\overline{v} \in V/U$. If $\overline{v} \notin B(V/U)$, then $\|v\|_V > 1$. So $v \notin B(V)$ for $v + u$ for all $u \in U$. Thus $f_{B(V)}(\overline{v}) = 0$ in this case. If $\overline{v} \in B(V/U)$, then $v + u \in B(V)$ for some $u \in U$. Thus $\|u\|_U \leq \|u + v\|_V + \|v\|_V \leq 2$. So $f_{B(V)}(\overline{v}) \leq \mu_U(2B(U)) = 2^{d_U} \cdot \mu_U(B(U))$ in this case. In either case, we have

$$\mu_V(B(V)) \leq 2^{d_U} \cdot \mu_U(B(U)) \cdot \mu_{V/U}(B(V/U)). \tag{2.2.5}$$

Next we turn to $f_{V/\Lambda_V}(\overline{v})$. Let $\{u_1, \ldots, u_{d_U}\}$ be a basis of $\Lambda_U = \Lambda_V \cap U$ and expand it to a basis $\{u_1, \ldots, u_{d_U}, v_1, \ldots, v_{d_V - d_U}\}$ of Λ_V. Then $\{\overline{v}_1, \ldots, \overline{v}_{d_V - d_U}\}$ is a basis of $\alpha(\Lambda_V)$. For each $\overline{v} \in (V/U)/\alpha(\Lambda_V)$, we have

$$f_{V/\Lambda_V}(\overline{v}) = \mu_U(\{u \in U : u + v \in V/\Lambda_V\}) = \mu_U(U/\Lambda_U).$$

Otherwise $f_{V/\Lambda_V}(\overline{v}) = 0$. So

$$\mu_V(V/\Lambda_V) = \mu_U(U/\Lambda_U) \cdot \mu_{V/U}((V/U)/\alpha(\Lambda_V)). \tag{2.2.6}$$

Now the conclusion follows from the definition of the volumes $\mathrm{Vol}(V) = \mu_V(B(V))/\mu_V(V/\Lambda_V)$ *etc.* □

2.3 Arakelov height of matrices

While the basic versions of Siegel's Lemma are sufficient for many applications, we state and prove a generalized version. Its proof, which is by the Geometry of Numbers and in particular uses the adelic version of Minkowski's second main theorem, is of particular importance.

Theorem 2.3.1. *Let A be an $M \times N$-matrix of rank M with entries in a number field K of degree d. Then the K-vector space of solutions of $A\mathbf{x} = \mathbf{0}$ has a basis $\mathbf{x}_1, \ldots, \mathbf{x}_{N-M}$, contained in $\mathcal{O}_K^N$, such that*

$$\prod_{l=1}^{N-M} H(\mathbf{x}_l) \leq |D_{K/\mathbb{Q}}|^{\frac{N-M}{2d}} H_{\mathrm{Ar}}(A),$$

where $D_{K/\mathbb{Q}}$ is the discriminant of K over $\mathbb{Q}$.

There are several things to be explained for this statement. First, $H(\mathbf{x}) = \exp(h(\mathbf{x}))$ is the multiplicative homogeneous height with $\mathbf{x}$ considered as a point in $\mathbb{P}^{N-1}(K)$; thus we may assume $\mathbf{x} \in \mathcal{O}_K^N$ because we can replace any solution by a non-zero scalar multiple and this does not change its height. Second, we need to define the *Arakelov height* $H_{\mathrm{Ar}}(A)$ of the matrix A; this is what we will do in this section.

Moreover, there is also a relative version for this generalized version. See Theorem 2.4.3.

2.3.1 Arakelov height on $\mathbb{P}^N$

Recall the Weil height which we defined before. For a point $\mathbf{x} = [x_0 : \cdots : x_N] \in \mathbb{P}^N(K)$, we have

$$[K : \mathbb{Q}]h(\mathbf{x}) = \sum_{v \in M_K^0} \log \max_j \|x_j\|_v + \sum_{v|\infty} \log \max_j \|x_j\|_v = \sum_{v \in M_K^0} \log \max_j \|x_j\|_v + \sum_{v|\infty} [K_v : \mathbb{R}] \log \max_j |x_j|_v.$$

There are other choices for the height function on $\mathbb{P}^N(\overline{\mathbb{Q}})$. In Arakelov theory, a more natural choice is to replace the L^∞-norm $\max_j |x_j|_v$ at the *archimedean* place by the L^2-norm $\left(\sum_{j=0}^N |x_j|_v^2\right)^{1/2}$. In other words, we define:

Definition 2.3.2. *For $\mathbf{x} = [x_0 : \cdots : x_N] \in \mathbb{P}^N(\overline{\mathbb{Q}})$ with each $x_j \in K$, define*

$$h_{\mathrm{Ar}}(\mathbf{x}) := \frac{1}{[K : \mathbb{Q}]} \left(\sum_{v \in M_K^0} \log \max_j \|x_j\|_v + \sum_{v|\infty} [K_v : \mathbb{R}] \log \left(\sum_{j=0}^N |x_j|_v^2 \right)^{1/2} \right).$$

One can check that $h_{\mathrm{Ar}}(\mathbf{x})$ is independent of the choice of the homogeneous coordinates (by the Product Formula) and of the choice of the number field K.

To ease notation, we introduce the following definition.

Definition 2.3.3. *For $\mathbf{x} = [x_0 : \ldots : x_N] \in \mathbb{P}^N(K)$ and $v \in M_K$, set*

$$H_v(\mathbf{x}) := \begin{cases} \max_j \|x_j\|_v = \max_j |x_j|_v^{[K_v : \mathbb{Q}_p]} & \text{if } v \text{ is non-archimedean,} \\ \left(\sum_{j=0}^N |x_j|_v^2\right)^{1/2 \cdot [K_v : \mathbb{R}]} & \text{if } v \text{ is archimedean.} \end{cases}$$

With this definition, the following holds true. For $\mathbf{x} = [x_0 : \cdots : x_N] \in \mathbb{P}^N(\overline{\mathbb{Q}})$ with each $x_j \in K$, we have

$$h_{\mathrm{Ar}}(\mathbf{x}) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in M_K} \log H_v(\mathbf{x}). \tag{2.3.1}$$

The following lemma will be proved in the Exercise class.

Lemma 2.3.4. *On $\mathbb{P}^N(\overline{\mathbb{Q}})$, the height functions h and h_{Ar} differ from a bounded function.*

Thus in view of the Height Machine, h_{Ar} is in the class represented by $h_{\mathbb{P}^N,\mathcal{O}(1)}$.

2.3.2 Height of matrices

We start by defining a height function on the Grassmannians. Let W be an M-dimensional subspace of $\overline{\mathbb{Q}}^N$. Then $\wedge^M W$ is a 1-dimensional subspace of $\wedge^M \overline{\mathbb{Q}}^N \simeq \overline{\mathbb{Q}}^{\binom{N}{M}}$. Thus we may view W as a point P_W of the projective space $\mathbb{P}(\wedge^M \overline{\mathbb{Q}}^N)$.

Definition 2.3.5. *The **Arakelov height** of W is defined to be $h_{\mathrm{Ar}}(W) := h_{\mathrm{Ar}}(P_W)$. We also define the multiplicative Arakelov height $H_{\mathrm{Ar}}(W) := \exp(h_{\mathrm{Ar}}(P_W))$.*

Now we are ready to define the Arakelov height of a matrix A.

Definition 2.3.6. *Let A be an $N \times M$-matrix with entries in $\overline{\mathbb{Q}}$.*

(i) Assume $\mathrm{rk}A = M$. Then $h_{\mathrm{Ar}}(A)$ is defined as $h_{\mathrm{Ar}}(W)$, where W is the subspace of $\overline{\mathbb{Q}}^N$ spanned by the columns of A.[4]

(ii) Assume $\mathrm{rk}A = N$. Then $h_{\mathrm{Ar}}(A) := h_{\mathrm{Ar}}(A^t)$ with A^t the transpose of A.

We also define the multiplicative Arakelov height $H_{\mathrm{Ar}}(A) := \exp(h_{\mathrm{Ar}}(P_W))$.

In general, A may not have the full rank. We then consider the subspace spanned by the columns or by the rows. This will lead to $h_{\mathrm{Ar}}^{\mathrm{col}}$ and $h_{\mathrm{Ar}}^{\mathrm{row}}$. We omit the definitions here but the idea will show up in the discussion of the generalized Siegel's Lemma in the next section.

We start with the following lemma, which makes the two parts of Definition 2.3.6 more "symmetric".

Lemma 2.3.7. *Let A be an $N \times M$-matrix with entries in $\overline{\mathbb{Q}}$. Assume $\mathrm{rk}A = N$. Then $h_{\mathrm{Ar}}(A)$ equals the Arakelov height of the subspace of $\overline{\mathbb{Q}}^M$ spanned by the rows of A.*

Proof. Consider the transpose A^t of A. It can be easily seen that A^t is an $M \times N$-matrix of rank N, and hence defines an injective linear map $\overline{\mathbb{Q}}^N \to \overline{\mathbb{Q}}^M$, which by abuse of notation we still denote by A^t. Part (i) of Definition 2.3.6 (applied to A^t) says that $h_{\mathrm{Ar}}(A^t)$ equals $h_{\mathrm{Ar}}(W)$ with $W \subseteq \overline{\mathbb{Q}}^M$ the subspace spanned by the columns of A^t. Notice that $W = \mathrm{Im}(A^t)$.

The matrix A defines a linear map $A\colon (\overline{\mathbb{Q}}^M)^* \to (\overline{\mathbb{Q}}^N)^*$ which is the dual of A^t. Consider the subspace $\mathrm{Ker}(A)$ of $(\overline{\mathbb{Q}}^M)^*$. Its annihilator $\mathrm{Ker}(A)^\perp$ in $((\overline{\mathbb{Q}}^M)^*)^* = \overline{\mathbb{Q}}^M$ then equals $\mathrm{Im}(A^t) = W$ by Linear Algebra. It is known that $\mathrm{Ker}(A)^\perp$ is spanned by the rows of A, and so is W. Hence we are done because $h_{\mathrm{Ar}}(A) = h_{\mathrm{Ar}}(A^t) = h_{\mathrm{Ar}}(W)$. □

Proposition 2.3.8. *Let W be an M-dimensional subspace of $\overline{\mathbb{Q}}^N$ and let $W^\perp$ be its annihilator in the dual $(\overline{\mathbb{Q}}^N)^* \simeq \overline{\mathbb{Q}}^N$. Then $h_{\mathrm{Ar}}(W^\perp) = h_{\mathrm{Ar}}(W)$.*

[4] Notice that A defines a linear map $A\colon \mathbb{R}^M \to \mathbb{R}^N$. The subspace W is precisely the image of this map. The assumption $\mathrm{rk}A = M$ is equivalent to the map A being injective.

This proposition has the following immediate corollary.

Corollary 2.3.9. *Let A be an $N \times M$-matrix with* $\mathrm{rk}A = N$ *and with entries in $\overline{\mathbb{Q}}$. Then the Arakelov height of the space of solutions of $A\mathbf{x} = \mathbf{0}$ equals $h_{\mathrm{Ar}}(A)$.*

Proof. We have $h_{\mathrm{Ar}}(A) = h_{\mathrm{Ar}}(A^t) = h_{\mathrm{Ar}}(\mathrm{Im}(A^t))$. But $\mathrm{Im}(A^t) = \mathrm{Ker}(A)^\perp$. So $h_{\mathrm{Ar}}(A) = h_{\mathrm{Ar}}(\mathrm{Ker}(A)^\perp)$, which then equals $h_{\mathrm{Ar}}(\mathrm{Ker}(A))$ by Proposition 2.3.8. Hence we are done. □

Proof of Proposition 2.3.8. Write $V = \overline{\mathbb{Q}}^N$. Any element $x \in \wedge^M V$ defines a linear map $\psi(x)\colon \wedge^{N-M} V \to \wedge^N V$, $y \mapsto x \wedge y$, and thus an element $\varphi(x) \in \wedge^N V \otimes \wedge^{N-M}(V^*)$. In other words, we obtained a map

$$\varphi\colon \wedge^M V \to \wedge^N V \otimes \wedge^{N-M}(V^*).$$

Then φ is an isomorphism and (better) each element of the canonical basis of $\wedge^M V$ is mapped to an element of the canonical basis of $\wedge^N V \otimes \wedge^{N-M}(V^*)$ up to a sign.

Notice that $\wedge^N V$ is a line. So it is easy to check that for any non-zero $x \in \wedge^M W$ (which is a line), the image of $\psi(x)$ is $\wedge^N V$ and the kernel of $\psi(x)$ is the subspace of $\wedge^{N-M} V$ generated by the elements of the form $w \wedge z$ with $w \in W$ and $z \in \wedge^{N-M-1} V$. Thus $\varphi(\wedge^M W) = \wedge^N V \otimes \wedge^{N-M}(W^\perp)$. Hence the coordinates of $\wedge^M W$ in $\mathbb{P}(\wedge^M V)$ are, up to a sign, equal to the coordinates of $\wedge^{N-M}(W^\perp)$ in $\mathbb{P}(\wedge^{N-M}(V^*))$. This proves the proposition. □

We finish this section by the following explicit formula for the definition of $h_{\mathrm{Ar}}(A)$. Let A be an $N \times M$-matrix with entries in $\overline{\mathbb{Q}}$.

For simplicity *we only consider the case* $\mathrm{rk}A = M$. Let $I \subseteq \{1, \dots, N\}$ with $|I| = M$. Denote by A_I the $M \times M$-submatrix of A formed with the i-th rows, $i \in I$, of A. Then the point in $\mathbb{P}(\wedge^N \overline{\mathbb{Q}}^M)$ corresponding to $\mathrm{Im}(A)$ is given by the coordinates $\det(A_I)$, where I ranges over all subsets of $\{1, \dots, N\}$ of cardinality M.

Let $K \subseteq \overline{\mathbb{Q}}$ be a number field which contains all entries of A. For each $v \in M_K$, set

$$H_v(A) := \begin{cases} \max_I |\det(A_I)|_v^{[K_v:\mathbb{Q}_p]} = \max_I \|\det(A_I)\|_v & \text{if } v \text{ is non-archimedean,} \\ \left(\sum_I |\det(A_I)|_v^2\right)^{1/2\cdot[K_v:\mathbb{R}]} = |\det(A^*A)|_v^{1/2\cdot[K_v:\mathbb{R}]} = \|\det(A^*A)\|_v^{1/2} & \text{if } v \text{ is archimedean.} \end{cases} \tag{2.3.2}$$

Here $A^* = \overline{A}^t$ is the adjoint of A, and $\sum_I |\det(A_I)|_v^2 = |\det(A^*A)|_v$ at the archimedean places by the *Binet Formula*.

Under this convention, we have

$$h_{\mathrm{Ar}}(A) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in M_K} \log H_v(A). \tag{2.3.3}$$

An immediate corollary of this explicit formula is:

Corollary 2.3.10. *Let G be an invertible $M \times M$-matrix. Then $h_{\mathrm{Ar}}(AG) = h_{\mathrm{Ar}}(A)$.*

Another application of this explicit formula is:

Corollary 2.3.11. *Let B and C be two complementary submatrices of A of type $N \times M_1$ and $M \times M_2$ respectively. Then $h_{\mathrm{Ar}}(A) \le h_{\mathrm{Ar}}(B) + h_{\mathrm{Ar}}(C)$.*

Proof. We only give a sketch. It suffices to prove $H_v(A) \le H_v(B)H_v(C)$ for each $v \in M_K$. If v is non-archimedean, it follows from Laplace's expansion. If v is archimedean, it follows from Fischer's inequality

$$\det \begin{pmatrix} B^*B & B^*C \\ C^*B & C^*C \end{pmatrix} \le \det(B^*B)\det(C^*C).$$

Alternatively, this corollary is an immediate consequence of the important theorem of Schmidt (independently of Struppeck–Vaaler) $h_{\mathrm{Ar}}(V+W) + h_{\mathrm{Ar}}(V \cap W) \le h_{\mathrm{Ar}}(V) + h_{\mathrm{Ar}}(W)$ for any subspaces V, W of $\overline{\mathbb{Q}}^M$. □

2.4 Generalized Siegel Lemma by Bombieri–Vaaler

The goal of this section is to have a deeper discussion of the generalized Siegel's Lemma by Bombieri–Vaaler (Theorem 2.3.1); in particular we give its proof. We repeat the statement here.

Theorem 2.4.1. *Let A be an $M \times N$-matrix of rank M with entries in a number field K of degree d. Then the K-vector space of solutions of $A\mathbf{x} = \mathbf{0}$ has a basis $\mathbf{x}_1, \dots, \mathbf{x}_{N-M}$, contained in $\mathcal{O}_K^N$, such that*

$$\prod_{l=1}^{N-M} H(\mathbf{x}_l) \leq |D_{K/\mathbb{Q}}|^{\frac{N-M}{2d}} H_{\mathrm{Ar}}(A),$$

where $D_{K/\mathbb{Q}}$ is the discriminant of K over $\mathbb{Q}$.

As said below Theorem 2.3.1, there is no deep information about the $\mathbf{x}_i$'s being contained in $\mathcal{O}_K^N$.

In practice, we may not always assume that A has maximal rank M. This can be obviated. We hereby state a corollary of Theorem 2.4.1, which bounds the heights of the solutions by the (multiplicative) Weil height instead of the Arakelov height.

Corollary 2.4.2. *Let A be an $M \times N$-matrix of rank R with entries in a number field K of degree d. Then there exists a basis $\mathbf{x}_1, \dots, \mathbf{x}_{N-R}$ of the kernel $\mathrm{Ker}(A)$, contained in $\mathcal{O}_K^N$, such that*

$$\prod_{l=1}^{N-R} H(\mathbf{x}_l) \leq |D_{K/\mathbb{Q}}|^{\frac{N-R}{2d}} \left(\sqrt{N} H(A)\right)^R.$$

Here $H(A)$ is the multiplicative Weil height of the point $[a_{ij}]_{i,j}$ viewed as a point in $\mathbb{P}^{MN-1}(K)$, with a_{ij} the entries of A.

In particular, there is a non-zero solution $\mathbf{x} \in \mathcal{O}_K^N$ of $A\mathbf{x} = \mathbf{0}$ with

$$H(\mathbf{x}) \leq |D_{K/\mathbb{Q}}|^{\frac{1}{2d}} \left(\sqrt{N} H(A)\right)^{\frac{R}{N-R}}.$$

2.4.1 Proof of Corollary 2.4.2 assuming Theorem 2.4.1

The "In particular" part follows clearly from the main part. So we will focus on proving the main part.

As $\mathrm{rk}A = R$, there is an $R \times N$-submatrix A' of A with $\mathrm{rk}A' = R$. Applying Theorem 2.4.1 to the matrix A', we get a basis $\mathbf{x}_1, \dots, \mathbf{x}_{N-R}$ of $\mathrm{Ker}(A)$ such that

$$\prod_{l=1}^{N-R} H(\mathbf{x}_l) \leq |D_{K/\mathbb{Q}}|^{\frac{N-R}{2d}} H_{\mathrm{Ar}}(A'). \tag{2.4.1}$$

On the other hand, if we denote by A_m the m-th row of A, then Corollary 2.3.11 implies that

$$H_{\mathrm{Ar}}(A') \leq \prod_m H_{\mathrm{Ar}}(A_m),$$

where m runs over the R rows of A'. Furthermore, the following inequality clearly holds true by definition

$$H_{\mathrm{Ar}}(A_m) \leq \sqrt{N} H(A).$$

Now, the two inequalities above yield $H_{\mathrm{Ar}}(A') \leq (\sqrt{N} H(A))^R$. So we can conclude by (2.4.1). □

2.4.2 Relative Version

As for Lemma 2.1.3 with respect to Lemma 2.1.1, we also have the following relative version of this generalized form of Siegel's Lemma.

Theorem 2.4.3. *Let K be a number field of degree d and F/K be a finite extension with $[F:K]=r$. Let A be an $M\times N$-matrix with entries in F.*

Assume $rM<N$. Then there exists $N-rM$ K-linearly independent vectors $\mathbf{x}_l\in\mathcal{O}_K^N$ such that $A\mathbf{x}_l=0$ for each $l\in\{1,\dots,N-rM\}$ and

$$\prod_{l=1}^{N-rM}H(\mathbf{x}_l)\le|D_{K/\mathbb{Q}}|^{\frac{N-rM}{2d}}\prod_{i=1}^{M}H_{\mathrm{Ar}}(A_i)^r,$$

where A_i is the i-th row of A.

The proof follows the guideline set up in Lemma 2.1.3.

Proof. Let $\omega_1,\dots,\omega_r$ be a basis of F/K. For the entries of $A=(a_{mn})$, we have

$$a_{mn}=\sum_{j=1}^{r}a_{mn}^{(j)}\omega_j$$

for uniquely determined $a_{mn}^{(j)}\in K$. Let $A^{(j)}$ be the $M\times N$-matrix with entries $a_{mn}^{(j)}$. Then for $\mathbf{x}\in K^N$, the equation $A\mathbf{x}=\mathbf{0}$ is equivalent to te system of equations $A^{(j)}\mathbf{x}=\mathbf{0}$ for all $j=1,\dots,r$. Write A' for the $rM\times N$-matrix $\begin{pmatrix}A^{(1)}\\ \vdots\\ A^{(r)}\end{pmatrix}$. Denote by $R:=\mathrm{rk}A'$.

It is attempting to apply Theorem 2.4.1 to A'. But we need to do one more step. Let $\sigma_1,\dots,\sigma_r$ be the distinct embeddings of F into $\overline{K}$ over K. Let Ω be the $rM\times rM$-matrix built up by r^2 blocks of $M\times M$-matrices $\Omega_{ij}=\sigma_i(\omega_j)I_M$. By construction of A', we have

$$A'':=\begin{pmatrix}\sigma_1A\\ \vdots\\ \sigma_rA\end{pmatrix}=\Omega A'.$$

From Algebraic Number Theory, it is known that $D_{F/K}=\det(\sigma_i(\omega_j))^2$. Thus Ω is invertible, and its inverse is again formed by r^2 blocks of multiples of I_M. In particular, $\mathrm{rk}A''=\mathrm{rk}A'=R$ and $\mathrm{Ker}(A'')=\mathrm{Ker}(A')$.

There exists an $R\times N$-submatrix A''' of A' with $\mathrm{rk}A'''=R$. Applying Theorem 2.4.1 to A''', we get a basis $\mathbf{x}_1,\dots,\mathbf{x}_{N-R}$ of $\mathrm{Ker}(A''')=\mathrm{Ker}(A')$, contained in $\mathcal{O}_K$, such that

$$\prod_{l=1}^{N-R}H(\mathbf{x}_l)\le|D_{K/\mathbb{Q}}|^{\frac{N-R}{2d}}H_{\mathrm{Ar}}(A'''),$$

If we denote by A_m the m-th row of A'', then Corollary 2.3.11 implies that

$$H_{\mathrm{Ar}}(A'')\le\prod_m H_{\mathrm{Ar}}(A_m''),$$

where m runs over the R rows of A''. Thus if we rearrange our basis $\mathbf{x}_l$ by increasing height, we have

$$\prod_{l=1}^{N-rM} H(\mathbf{x}_l) \le \left(\prod_{l=1}^{N-R} H(\mathbf{x}_l)\right)^{\frac{N-rM}{N-R}} \le |D_{K/\mathbb{Q}}|^{\frac{N-rM}{2d}} \left(\prod_m H_{\mathrm{Ar}}(A''_m)\right)^{\frac{N-rM}{N-R}}. \tag{2.4.2}$$

By definition of the Arakelov height, we have H_{Ar} takes value in $[1,\infty)$. Thus $(\prod_m H_{\mathrm{Ar}}(A''_m))^{\frac{N-rM}{N-R}} \le \prod_{i=1}^{rM} H_{\mathrm{Ar}}(A''_i)$. Now the conclusion follows because H_{Ar} is invariant under each σ_i. □

Below is reading material. It will not be covered in the main lectures or in the Exercise class.

2.4.3 Adelic version of Minkowski's Second Theorem

The proof of Theorem 2.4.1 uses geometry of numbers over the adèles and Minkowski's Second Theorem. In this subsection, we introduce/recall these prerequisites.

Let K be a number field, $v \in M_K$ and K_v be the completion of K with respect to v. It is known that K_v is a locally compact group.

The **ring of adèles** of K is the subring

$$\mathbb{A}_K := \{\mathbf{x} = (x_v) \in \prod_{v\in M_K} K_v : x_v \in R_v \text{ up to finitely many } v\}.$$

of $\prod_{v\in M_K} K_v$.

One should be careful with the *topology* on $\mathbb{A}_K$. It is *not* induced by the product topology on $\prod_{v\in M_K} K_v$! Rather, we consider for each finite subset $S \subseteq M_K$ containing all archimedean places the product

$$H_S := \prod_{v\in S} K_v \times \prod_{v\notin S} R_v.$$

The product topology makes each such H_S into a locally compact topological group. The topology which we put on $\mathbb{A}_K$ is *the unique topology such that the groups H_S are open topological subgroups of $\mathbb{A}_K$*. In fact, this makes $\mathbb{A}_K$ a locally compact topological ring.

It is known that the diagonal map $K \to \mathbb{A}_K$, $x \mapsto (x_v)_{v\in M_K}$, makes K into a *discrete closed subgroup* of $\mathbb{A}_K$. Moreover $\mathbb{A}_K/K$ is compact.

Let $v|p \in M_{\mathbb{Q}}$. Then K_v is a locally compact group with Haar measure uniquely determined up to a scalar. We normalize this Haar measure as follows:

(a) if v is non-archimedean, β_v denotes the Haar measure on K_v normalized so that

$$\beta_v(R_v) = |D_{K_v/\mathbb{Q}_p}|_p^{1/2}$$

where R_v is the valuation ring of K_v and $D_{K_v/\mathbb{Q}_p}$ is the discriminant;

(b) if $K_v = \mathbb{R}$, then β_v is the usual Lebesgue measure;

(c) if $K_v = \mathbb{C}$, then β_v is twice the usual Lebesgue measure.

For each finite subset $S \subseteq M_K$ containing all archimedean places, the product measure $\beta_S := \prod_{v\in S} \beta_v \times \prod_{v\notin S} \beta_v|_{R_v}$ is then a Haar measure on the open topological subgruop H_S of $\mathbb{A}_K$. The measures β_S fit together to give a Haar measure β on $\mathbb{A}_K$.[5]

Let N be a positive integer. For each (archimedean) $v|\infty$, let S_v be a non-empty convex, symmetric, open subset of K_v^N; here "symmetric" means $S_v = -S_v$. For each (non-archimedean) $v \in M_K^0$, let S_v be

[5] With this in hand, we can shortly explain why we take the normalizations above. The Haar measure β on $\mathbb{A}_K$ induces a Haar measure $\beta_{\mathbb{A}_K/K}$ on the compact group $\mathbb{A}_K/K$, and the normalization above makes the volume of $\mathbb{A}_K/K$ to be 1.

a K_v-lattice in K_v^N, *i.e.* a non-empty compact open R_v-submodule of K_v^N. *Assume that* $S_v = R_v^N$ *for all but finitely many* v. Then the set

$$\Lambda := \{\mathbf{x} \in K^N : \mathbf{x} \in S_v \text{ for all } v \in M_K^0\}$$

is a K-lattice in K^N, *i.e.* a finitely generated $\mathcal{O}_K$-module which generates K^N as a vector space. Moreover, the image Λ_∞ of Λ under the canonical embedding $K^N \hookrightarrow E_\infty := \prod_{v|\infty} K_v^N$ is an $\mathbb{R}$-lattice in E_∞.[6]

Definition 2.4.4. *The* n***-th successive minimum*** *of the non-empty convex symmetric open subset* $S_\infty := \prod_{v|\infty} S_v$ *of* E_∞ *with respect to the lattice* Λ_∞ *is*

$$\lambda_n := \inf\{t > 0 : tS_\infty \text{ contains } n \text{ } K\text{-linearly independent vectors of } \Lambda_\infty\}.$$

Now we are ready to state (the adelic version of) Minkowski's Second Theorem.

Theorem 2.4.5 (Minkowski's Second Theorem, adelic form)**.** *The successive minima defined above satisfy*

$$(\lambda_1 \cdots \lambda_N)^d \prod_{v \in M_K} \beta_v(S_v) \le 2^{dN}.$$

Here, the product $\prod_{v\in M_K} \beta_v(S_v)$ should be understood to be the *volume* of S with respect to the Haar measure on $\mathbb{A}_K$ defined by the β_v's at each $v \in M_K$.

2.4.4 Setup for the application of Minkowski's Second Theorem

For the purpose of proving Siegel's Lemma in the form of Theorem 2.4.1, we do the following preparation.

For the sets S_v: First, let Q_v^N be the unit cube in K_v^N of volume 1 with respect to the Haar measure β_v. More explicitly, $\mathbf{x} = (x_1, \dots, x_N) \in Q_v^N$ if and only if

$$\begin{cases} \max_n \|x_n\|_v < \frac{1}{2} & \text{if } v \text{ is real} \\ \max_n \|x_n\|_v < \frac{1}{2\pi} & \text{if } v \text{ is complex} \\ \max_n \|x_n\|_v \le 1 & \text{if } v \text{ is non-archimedean.} \end{cases}$$

Let A be an $N \times M$-matrix with entries in K such that $\mathrm{rk} A = M$. Set

$$S_v := \{\mathbf{y} \in K_v^M : A\mathbf{y} \in Q_v^N\}. \tag{2.4.3}$$

If v is archimedean, then S_v is a non-empty convex symmetric bounded open subset of K_v^N; indeed, under the injective linear map $\mathbf{x} \mapsto A\mathbf{x}$, the image of S_v is a linear slice of the cube Q_v^N. If v is non-archimedean, then one can show that S_v is a K_v-lattice in K_v^M and that $S_v = R_v^M$ for all but finitely many v; in fact in this case we have the following more precise result.

Proposition 2.4.6. *Let* $v \in M_K^0$ *lying over the prime number* p. *Then* S_v *is a* K_v*-lattice in* K_v^M *and* $S_v = R_v^M$ *for all but finitely many* v. *Moreover, we have*

$$\beta_v(S_v) = |D_{K_v/\mathbb{Q}_p}|_p^{M/2} \left(\max_I \|\det(A_I)\|_v\right)^{-1},$$

where I *runs over all subsets of* $\{1, \dots, N\}$ *of cardinality* M, *and* A_I *is the* $M \times M$*-matrix formed by the* i*-th rows of* A *with* $i \in I$.

[6] This is the familiar notion of a lattice, namely Λ_∞ is a discrete subgroup fo the $\mathbb{R}$-vector space E_∞ and that E_∞/Λ_∞ is compact.

Proof. Choose a subset $J \subseteq \{1, \dots, N\}$ of cardinality M such that $\| \det(A_J) \|_v = \max_I \| \det(A_I) \|_v$. Without loss of generality, we may assume $J = \{1, \dots, M\}$. Then $W := AA_J^{-1}$ is of the form

$$W = \begin{pmatrix} I_M \\ W' \end{pmatrix}.$$

For any subset $I \subseteq \{1, \dots, N\}$ of cardinality M, we have $\| \det(W_I) \|_v \leq 1$ by choice of J. In particular, taking $I = \{1, \dots, l-1, l+1, \dots, M, M+j\}$ we get

$$\|w_{M+j,l}\|_v = \| \det(W_I) \|_v \leq 1.$$

Thus all entries of W are in the valuation ring R_v and this proves

$$A_J S_v = \{\mathbf{y} \in K_v^M : W\mathbf{y} \in Q_v^N\} = R_v^M. \tag{2.4.4}$$

This proves that S_v is a K_v-lattice in K_v^M and that $S_v = R_v^M$ for all but finitely many v

It remains to compute $\beta_v(S_v)$. It is known that under the linear transformation $\mathbf{y} \mapsto A_J^{-1}\mathbf{y}$ on K_v^M, the volume transforms by the factor $\| \det(A_J) \|_v^{-1}$. Thus

$$\beta_v(S_v) = \| \det(A_J) \|_v^{-1} \beta_v(R_v^M) = \| \det(A_J) \|_v^{-1} |D_{K_v/\mathbb{Q}_p}|_p^{M/2}$$

which is what we desire. □

We also need to bound $\beta_v(S_v)$ from below for v archimedean. For this purpose, we have

Proposition 2.4.7. *Let $v \in M_K$ with $v|\infty$. Then*

$$\beta_v(S_v) \geq \| \det(A^*A) \|_v^{-1/2}$$

where $A^* = \overline{A}^t$ is the adjoint of A.

Proof. The proof uses *Vaaler's cube-slicing theorem*, which we state here without proof.

Vaaler's cube-slicing theorem. Let $N = n_1 + \cdots + n_r$ be a partition. Let $Q_N := B_{\rho(n_1)} \times \cdots \times B_{\rho(n_r)}$, where each $B_{\rho(n_j)}$ is the closed ball of volume 1 in $\mathbb{R}^{n_j}$ centered at 0.[7] For a real $N \times M$-matrix B of rank M, we have

$$\det(B^t B)^{-1/2} \leq \mathrm{Vol}\left(\{\mathbf{y} \in \mathbb{R}^M : B\mathbf{y} \in Q_N\}\right). \tag{2.4.5}$$

An easier way to understand this volume bound is as follows. Let $L := \mathrm{Im}(B) \subseteq \mathbb{R}^N$ which is an M-dimensional subspace. Then (2.4.5) is equivalent to $1 \leq \mathrm{Vol}(Q_N \cap L)$, *i.e.* the volume of a slice through the center of a product of balls of volume 1 is bounded below by 1.

Now we go back to the proof of Proposition 2.4.7. If $K_v = \mathbb{R}$, then this is (2.4.5) for $r = N$ and $n_1 = \cdots = n_N = 1$. Assume $K_v = \mathbb{C}$. Write $A = U + \sqrt{-1}V$ and $\mathbf{y} = \mathbf{u} + \sqrt{-1}\mathbf{v}$ for real $U, V, \mathbf{u}, \mathbf{v}$. Thus $K_v^M \simeq \mathbb{R}^{2M}$, $\mathbf{y} \mapsto (\mathbf{u}, \mathbf{v})$. Similarly we have $K_v^N \simeq \mathbb{R}^{2N}$. Now, the linear map $\mathbf{y} \mapsto A\mathbf{y}$ is given by the real $2N \times 2M$-matrix

$$A' = \begin{pmatrix} U & -V \\ V & U \end{pmatrix}$$

and

$$Q_v^N = \left\{(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^{2N} : u_j^2 + v_j^2 < \frac{1}{2\pi}\right\}.$$

By (2.4.5) for $n_1 = \cdots = n_N = 2$, we then have

$$\beta_v(S_v) \geq \det(A'^t A')^{-1/2}.$$

Since $A \mapsto A'$ is a ring homomorphism from the complex $N \times M$-matrices to the real $2N \times 2M$-matrices, we have $\det(A'^t A') = \det((A^*A)') = \det(A^*A)^2$. Hence we can conclude. □

[7] So the radius of $B_{\rho(n_j)}$ is $\rho(n_j) = \pi^{-1/2}\Gamma(n_j/2+1)^{1/n_j}$.

2.4.5 Proof of Theorem 2.4.1

With the preparation from last subsection, we prove Bombieri–Vaaler's Siegel Lemma in this subsection. We start with:

Proposition 2.4.8. *Let A be an $N \times M$-matrix of rank M with entries in K. Then the image of A has a basis $\mathbf{x}_1, \dots, \mathbf{x}_M$ with*

$$\prod_{m=1}^{M} H(\mathbf{x}_m) \le \left(\frac{2}{\pi}\right)^{\frac{Ms}{d}} |D_{K/\mathbb{Q}}|^{\frac{M}{2d}} H_{\mathrm{Ar}}(A)$$

where s is the number of complex places of K and $d = [K : \mathbb{Q}]$.

Proof. By Proposition 2.4.6 and Proposition 2.4.7, we have

$$\prod_{v \in M_K} \beta_v(S_V) \ge \prod_{v \in M_K^0} |D_{K_v/\mathbb{Q}_p}|_p^{M/2} \left(\prod_{v \in M_K^0} \max_I \| \det(A_I) \|_v \cdot \prod_{v \mid \infty} \| \det(A^* A) \|_v^{1/2} \right)^{-1}.$$

By (2.3.3), this becomes

$$\prod_{v \in M_K} \beta_v(S_V) \ge \left(\prod_{v \in M_K^0} |D_{K_v/\mathbb{Q}_p}|_p^{M/2} \right) H_{\mathrm{Ar}}(A)^{-d}.$$

It is known, from Algebraic Number Theory, that $|D_{K/\mathbb{Q}}|_p = \prod_{v|p} |D_{K_v/\mathbb{Q}_p}|_p$ for each prime number p. Thus the Product Formula implies $|D_{K/\mathbb{Q}}|^{-1} = \prod_{v \in M_K^0} |D_{K_v/\mathbb{Q}_p}|_p$. So the inequality above becomes

$$\prod_{v \in M_K} \beta_v(S_V) \ge |D_{K/\mathbb{Q}}|^{-M/2} H_{\mathrm{Ar}}(A)^{-d}.$$

Thus, Minkowski's Second Theorem, Theorem 2.4.5, yields

$$\lambda_1 \cdots \lambda_M \le 2^M |D_{K/\mathbb{Q}}|^{M/2d} H_{\mathrm{Ar}}(A). \tag{2.4.6}$$

It remains to use the successive minima find the desired basis. For the specific sets S_v constructed in (2.4.3), recall the K-lattice $\Lambda = \{\mathbf{x} \in K^N : \mathbf{x} \in S_v \text{ for all } v \in M_K^0\}$ which is identified with its image Λ_∞ under the canonical embedding $K^N \hookrightarrow E_\infty = \prod_{v|\infty} K_v^N$. Let $\mathbf{y} \in K^M$ be a lattice point in λS_∞ for some $\lambda > 0$ and let $\mathbf{x} = A\mathbf{y}$. Then the definition of $S_\infty = \prod_{v|\infty} S_v$ yields $\max_n \|x_n\|_v < \lambda/2$ if v is real, $\max_n \|x_n\|_v < \lambda^2/2\pi$ if v is complex, and $\max_n \|x_n\|_v \le 1$ if $v \in M_K^0$. Thus we have

$$H(A\mathbf{y}) < \frac{\lambda}{2} \left(\frac{2}{\pi}\right)^{s/d}. \tag{2.4.7}$$

By the definition of successive minima, there are linearly independent lattice points $\mathbf{y}_1, \dots, \mathbf{y}_M \in K^M$ such that $\mathbf{y}_m \in \lambda_m S_\infty$ for each $m \in \{1, \dots, M\}$. Then we obtain the desired basis from (2.4.6) and (2.4.7), with $\mathbf{x}_m = A\mathbf{y}_m$. □

Proof of Theorem 2.4.1. For the $M \times N$-matrix A of rank M, its transpose A^t is an $N \times M$-matrix of rank M. It is attempting to apply Proposition 2.4.8 directly to A^t, but we need to do more.

We wish to find a basis of $\mathrm{Ker}(A)$ of small height. To do this, we first of all take an arbitrary basis $\mathbf{y}_1, \dots, \mathbf{y}_{N-M}$ of $\mathrm{Ker}(A)$, and let $A' := \begin{pmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_M \end{pmatrix}$. Then A' is an $N \times (N-M)$-matrix with rank $N - M$, and $\mathrm{Im}(A') = \mathrm{Ker}(A)$. Recall that $h_{\mathrm{Ar}}(A) = h_{\mathrm{Ar}}(\mathrm{Ker}(A))$ by Corollary 2.3.9. Hence $h_{\mathrm{Ar}}(A') = h_{\mathrm{Ar}}(A)$.

Apply Proposition 2.4.8 to A'. Then we get a basis $\mathbf{x}_1, \dots, \mathbf{x}_{N-M}$ of $\mathrm{Im}(A') = \mathrm{Ker}(A)$ such that

$$\prod_{l=1}^{N-M} H(\mathbf{x}_l) \le \left(\frac{2}{\pi}\right)^{(N-M)s/d} |D_{K/\mathbb{Q}}|^{\frac{N-M}{2d}} H_{\mathrm{Ar}}(A').$$

But $2/\pi < 1$. So we are done because $H_{\mathrm{Ar}}(A') = H_{\mathrm{Ar}}(A)$. □

Chapter 3

Roth's Theorem

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

3.1.1 From Liouville to Thue

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

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

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

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

Proof. We will divide the proof into several steps.
Step I: Construct an auxiliary polynomial Let $f(x) \in \mathbb{Z}[x]$ be the minimal polynomial of α over $\mathbb{Q}$ with relatively prime integral coefficients. In particular, f is irreducible over $\mathbb{Q}$ and has degree d.
Step II: Non-vanishing at the rational point If $p/q \in \mathbb{Q}$, then we have $f(p/q) \neq 0$.
Step III: Lower bound (Liouville) By Step II, we then have $|f(p/q)| \geq 1/q^d$ since $\deg f = d$.
Step IV: Upper bound As $f(\alpha) = 0$ and f is the minimal polynomial of α, we can write $f(x) = (x - \alpha)g(x)$ with $g(\alpha) \neq 0$. Thus

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.1.2 Statement of Roth's Theorem

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

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

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

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

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

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

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

3.2 Index and preparation of the construction of the auxiliary polynomial

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

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

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

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

and

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

Then

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

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

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

Now let us define the index.

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

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

Lemma 3.2.3. *The following properties hold true.*

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

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

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

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

For (ii): Assume that $\operatorname{ind}(PQ; \mathbf{d}; \boldsymbol{\alpha})$ is achieved at some $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$. We have $\partial_{\mu}(PQ) = \sum_{\boldsymbol{\mu_1} + \boldsymbol{\mu_2} = \boldsymbol{\mu}} C_{\boldsymbol{\mu_1}, \boldsymbol{\mu_2}} (\partial_{\boldsymbol{\mu_1}} P)(\partial_{\boldsymbol{\mu_2}} Q)$ for some positive integers $C_{\boldsymbol{\mu_1}, \boldsymbol{\mu_2}}$.[2] Thus there exists $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ such that $\boldsymbol{\mu_1} + \boldsymbol{\mu_2} = \boldsymbol{\mu}$, $\partial_{\boldsymbol{\mu_1}} P(\boldsymbol{\alpha}) \neq 0$ and $\partial_{\boldsymbol{\mu_2}} Q(\boldsymbol{\alpha}) \neq 0$. Thus the definition of index yields $\sum_j \frac{\mu_{1,j}}{d_j} \geq \operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha})$ and $\sum_j \frac{\mu_{2,j}}{d_j} \geq \operatorname{ind}(Q; \mathbf{d}; \boldsymbol{\alpha})$. So $\operatorname{ind}(PQ; \mathbf{d}; \boldsymbol{\alpha}) = \sum_j \frac{\mu_{1,j} + \mu_{2,j}}{d_j} \geq \operatorname{ind}(P; \mathbf{d}; \boldsymbol{\alpha}) + \operatorname{ind}(Q; \mathbf{d}; \boldsymbol{\alpha})$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The number M of equations is $M = \#(\Gamma \cap \mathcal{V}_m(t))$ for the lattice $\Gamma = \frac{1}{d_1}\mathbb{Z} \times \cdots \frac{1}{d_m}\mathbb{Z}$. We claim that $M \sim V_m(t) d_1 \cdots d_m$ as $d_j \to \infty$ for all $j \in \{1, \dots, m\}$. Indeed, $V_m(t) d_1 \cdots d_m \leq M$

because we can associate to each lattice point in Γ the parallelopiped $[i_1/d_1, (i_1+1)/d_1] \times \cdots \times [i_m/d_m, (i_m+1)/d_m]$. On the other hand, for each $(i_1/d_1, \ldots, i_m/d_m) \in \Gamma \cap \mathcal{V}_m(t)$, we have

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

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

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

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

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

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

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

Next we give an estimate of the volume in question.

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

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

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

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

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

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

3.3 Proof of Roth's Theorem assuming zero estimates

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

We start by restating Roth's Theorem.

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

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

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

3.3.0 Step 0: Choosing independent solutions.

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

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

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

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

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

3.3.1 Step 1: Construction of an auxiliary polynomial.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.3.3 Step 3: Lower bound (Liouville).

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

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

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

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

3.3.4 Step 4: Upper bound.

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

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

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

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

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

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

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

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

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

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

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

3.3.5 Step 5: Comparison of the two bounds.

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

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

Dividing both sides by mD, we get

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

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

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

3.4 Zero estimates: Roth's Lemma

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

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

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

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

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

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

Then we have

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

3.4.1 Proof of Lemma 3.3.2 by Roth's Lemma

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

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

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

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

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

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

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

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

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

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

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

3.4.2 Proof of Roth's Lemma

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

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

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

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

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

Thus

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

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

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

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

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

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

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

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

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

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

Set

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

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

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

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

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

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

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

For heights, Lemma 1.3.3 yields $h(W) = h(U) + h(V)$. We claim that

$$h(W) \leq (s+1)(h(P) + 4d_1). \tag{3.4.4}$$

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

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

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

Since $d_{j+1}/d_j \le \sigma \le 1/2$ by hypothesis (i), we have $d_1 + \ldots + d_m \le 2d_1$. On the other hand, $\log(s+1)! \le (s+1)\log(s+1) \le (s+1)\log(d_m+1) \le (s+1)d_m \le (s+1)d_1/2$. Hence we can establish (3.4.4).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence we are done. □

3.4.3 Proof of Proposition 3.4.2

We start with $\Leftarrow$. Assume $\varphi_1, \ldots, \varphi_n$ are linearly dependent over $\overline{\mathbb{Q}}$. Then all generalized Wronskians vanish. Indeed, we have $c_1\varphi_1 + \cdots + c_n\varphi_n = 0$ for some $c_1, \ldots, c_n \in \overline{\mathbb{Q}}$ not all zero. Applying the operators $\partial_{\boldsymbol{\mu}_i}$ to this relation, we obtain a linear system in the coefficients c_j and its determinant mush vanish. This determinant is precisely the generalized Wronskian $W_{\boldsymbol{\mu_1},\ldots,\boldsymbol{\mu_n}}(x_1, \ldots, x_m)$.

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

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

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

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

is not identically zero.

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

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

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

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

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

is not identically 0. But

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

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

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

Assume $\Rightarrow$ is proved for $1, \ldots, n-1$. For n and the polynomials $f_1, \ldots, f_n$, assume that $W(t)$ is identically 0. For each $j \in \{1, \ldots, n\}$, set $W_j(t)$ to be the Wronskian of the $n-1$ polynomials by omitting

f_j. Then by expanding the determinant $W(t)$ by the last row, we get $W(t) = \sum_{j=1}^{n} W_j \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n-1} f_j = \sum_{j=1}^{n} W_j f_j^{(n-1)}$. Here we change the notation and denote by $f_j^{(i)}$ the i-th derivative of f_j. Thus

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

We claim that $W_1 f_1 + \cdots + W_n f_n \equiv 0$. Indeed, the left hand side is the determinant of the $n \times n$-matrix $\begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ \cdot & \cdot & \cdots & \cdot \\ f_1^{(n-2)} & f_2^{(n-2)} & \cdots & f_n^{(n-2)} \\ f_1 & f_2 & \cdots & f_n \end{pmatrix}$, by the expansion along the last row. Similarly we have $\sum_j W_j f_j^{(i)} \equiv 0$ for each $i \in \{1, \ldots, n-2\}$. Thus we obtain a system of n equalities of polynomials

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

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

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

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

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

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

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

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

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

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

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

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

In the proof of Roth's Theorem presented in previous sections of this chapter, we used Roth's Lemma (Lemma 3.4.1) to do the zero estimates and found a polynomial P having large index at $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)$ but small index at $(p_1/q_1, \ldots, p_m/q_m)$. Roth's Lemma is *arithmetic* in nature:

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

the polynomial P has coefficients in $\overline{\mathbb{Q}}$, we are interested in its order of vanishing at an algebraic point, and a hypothesis (hypothesis (ii)) on the given data is about the heights.

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

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

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

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

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

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

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

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

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

We need the following technical lemma.

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

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

Proof. We prove the lemma by taking τ such that

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

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

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

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

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

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

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

So

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Comparing (3.5.5) and (3.5.6), we get

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

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

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

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

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

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

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

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

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

Chapter 4

Abelian Varieties

4.1 Algebraic curves

The goal of this section is to gather some results for algebraic curves. Below, by k we always mean an algebraic closed field of characteristic 0. Typical examples are $k = \mathbb{C}$ or $k = \overline{\mathbb{Q}}$.

For simplicity, by a *curve*, we always mean an irreducible projective variety of dimension 1 defined over k.

4.1.1 Basic definitions

Definition 4.1.1. *Let $V \subseteq \mathbb{A}^n$ be a variety, $P \in V$, and assume $I(V) = (f_1, \ldots, f_m)$. Then the rank of the $m \times n$ matrix*

$$(\partial f_i / \partial X_j(P))_{1 \leq i \leq m, 1 \leq j \leq n}$$

*is constant outside a proper Zariski closed subset of V. The **dimension of** V, denoted by $\dim V$, is defined to be n minus this rank.*

*We say that V is **non-singular** (or **smooth**) at P if the rank of the matrix above evaluated at P is $n - \dim V$.*

This definition is local. If $V \subseteq \mathbb{P}^n$ and $P \in V$, then to define the smoothness of V at P we take any affine chart in which P lies. We also say that V is **non-singular** (or **smooth**) if V is smooth at every point.

Definition-Proposition 4.1.2. *Let C be a curve and $P \in C$ a smooth point. Then $\mathcal{O}_{C,P}$ is a discrete valuation ring.*

*Denote by $\mathfrak{m}_P$ the maximal ideal of $\mathcal{O}_{C,P}$. The **(normalized) valuation** on $\mathcal{O}_{C,P}$ is given by*

$$\begin{array}{rccc} \mathrm{ord}_P\colon & \mathcal{O}_{C,P} & \to & \{0,1,2,\ldots\} \cup \{\infty\} \\ & f & \mapsto & \max\{d \in \mathbb{Z} : f \in \mathfrak{m}_P^d\}. \end{array}$$

Using $\mathrm{ord}_P(f/g) = \mathrm{ord}_P(f) - \mathrm{ord}_P(g)$, we extend this function ord_P to $k(C)$ (the field of rational functions on C)

$$\mathrm{ord}_P\colon k(C) \to \mathbb{Z} \cup \{\infty\}. \tag{4.1.1}$$

*A **uniformizer for** C **at** P is a function $t \in k(C)$ with $\mathrm{ord}_P(t) = 1$, i.e. a generator of $\mathfrak{m}_P$.*

Example 4.1.3. *Suppose that $K = \mathbb{C}$ and $f \in k(C)$ is a meromorphic function on C. Then $\mathrm{ord}_P(f) = 0$ if and only if P is neither a zero nor a pole of f. On the other hand, if P is a zero of f, then $\mathrm{ord}_P(f) > 0$; if P is a pole of f, then $\mathrm{ord}_P(f) < 0$.*

More concretely, consider the curve $C \subseteq \mathbb{P}^2$ whose intersection with $\mathbb{A}^2$ is

$$y^2 = x^3 - x = x(x+1)(x-1).$$

Let $P = (0,0)$. Then C is smooth at P. It is not hard to see that $\mathfrak{m}_P$ is generated by x, y and $\mathfrak{m}_P^2$ is generated by x^2, xy, y^2. Thus $x = x^3 - y^2 \equiv 0 \pmod{\mathfrak{m}_P^2}$.[1] *One can check for example*

$$\mathrm{ord}_P(y) = 1,\ \mathrm{ord}_P(x) = 2,\ \mathrm{ord}_P(2y^2 - x) = 2.$$

4.1.2 Divisors

Definition 4.1.4. *A **divisor** D on a curve C is a formal sum*

$$D = \sum_{P \in C} n_P [P]$$

*with $n_P \in \mathbb{Z}$ and $n_P = 0$ for all but finitely many $P \in C$. The **degree of** D is defined to be* $\deg D := \sum_{P \in C} n_P$.

*The **support** of D is $\{P \in C : n_P \neq 0\}$.*

The set of all divisors on C is denoted by $\mathrm{Div}(C)$*; it is a free abelian group (which is generated by the points of C). We will call it the **divisor group of** C.*

A particular type of divisors on C is constructed in the following way. Assume C is smooth and let $f \in k(C)^*$. Then we can associated with f a divisor

$$\mathrm{div}(f) := \sum_{P \in C} \mathrm{ord}_P(f)[P].$$

Proposition 4.1.5. *Let C be a smooth curve and let $f \in k(C)^*$. Then* $\deg \mathrm{div}(f) = 0$*. Moreover,* $\mathrm{div}(f) = 0$ *if and only if $f \in k^*$.*

Definition 4.1.6. *A divisor $D \in \mathrm{Div}(C)$ is **principal** if it equals* $\mathrm{div}(f)$ *for some $f \in k(C)^*$.*

*Two divisors D_1, D_2 are **linearly equivalent**, denoted $D_1 \sim D_2$, if $D_1 - D_2$ is a principal divisor.*

In particular, linearly equivalent divisors have the same degree. We define the ***divisor class group of*** C to be $\mathrm{Cl}(C) := \mathrm{Div}(C)/\sim$.

Example 4.1.7. *Consider the curve in $C \subseteq \mathbb{P}^2$ whose intersection with $\mathbb{A}^2$ is*

$$y^2 = x^3 - x = x(x+1)(x-1).$$

One can check that C is smooth and it has a single point at infinity, which we denote by ∞. Set $P_1 = (0,0)$, $P_2 = (1,0)$ and $P_3 = (-1,0)$. Then (see the end of Example 4.1.3)

$$\mathrm{div}(x) = 2[P_1] - 2[\infty] \quad \textit{and} \quad \mathrm{div}(Y) = [P_1] + [P_2] + [P_3] - 3[\infty].$$

A partial order on $\mathrm{Div}(C)$ can be put as follows.

Definition 4.1.8. *A divisor $D = \sum n_P [P] \in \mathrm{Div}(C)$ is **effective**, denoted by $D \geq 0$, if $n_P \geq 0$ for every $P \in C$.*

If $D_1, D_2 \in \mathrm{Div}(C)$, then we write $D_1 \geq D_2$ if $D_1 - D_2 \geq 0$.

[1] In order words, $\mathfrak{m}_P/\mathfrak{m}_P^2$ is generated by $\overline{y}$.

Remark 4.1.9. *Divisorial inequalities are a useful tool for describing poles and zeros of functions. Let us see the following example.*

Let $f \in k(C)^*$ be a function which has a pole of order at most n at a point $P \in C$ and is regular everywhere else. This requirement on f is equivalent to $\operatorname{div}(f) \geq -n[P]$.

If furthermore we require f to have a zero at a point $Q \in C$, then the requirement becomes $\operatorname{div}(f) \geq [Q] - n[P]$.

Definition-Proposition 4.1.10. *Let $D \in \operatorname{Div}(C)$. We associate to D the set of functions*

$$L(D) := \{f \in k(C)^* : \operatorname{div}(f) \geq -D\} \cup \{0\}.$$

Then $L(D)$ is a finite-dimensional k-vector space. Denote its dimension by

$$\ell(D) := \dim_k L(D). \tag{4.1.2}$$

Lemma 4.1.11. *Let $D \in \operatorname{Div}(C)$.*

(i) If $\deg(D) < 0$, then $\ell(D) = 0$.

(ii) $\ell(0) = 1$.

(iii) If $D' \in \operatorname{Div}(C)$ is linearly equivalent to D, then $L(D) \simeq L(D')$ and so $\ell(D) = \ell(D')$.

Proof. For (i): Suppose there exists $0 \neq f \in L(D)$. Then $0 = \deg \operatorname{div}(f) \geq \deg(-D) = -\deg D$. Hence $\deg D \geq 0$.

For (ii): For each $f \in k(C)^*$, we have $f \in L(0) \Leftrightarrow \operatorname{div}(f) \geq 0$. So $f \in L(0)$ if and only if f is a regular function on C. But C is projective, so f must be a constant. So $\dim_k L(0) = 1$.

For (iii): Suppose $D' = D + \operatorname{div}(g)$ for $g \in k(C)^*$. Then the map $L(D') \to L(D)$, $f \mapsto fg$, is an isomorphism. □

4.1.3 Differentials and canonical divisor

Definition 4.1.12. *The **space of (meromorphic) differential forms** on C, denoted by Ω_C, is the $k(C)$-vector space generated by symbols of the form $\mathrm{d}x$ for $x \in k(C)$ satisfying the following properties:*

(i) $\mathrm{d}(x+y) = \mathrm{d}x + \mathrm{d}y$ for all $x, y \in k(C)$;

(ii) $\mathrm{d}(xy) = x\mathrm{d}y + y\mathrm{d}x$ for all $x, y \in k(C)$;

(ii) $\mathrm{d}a = 0$ for all $a \in K$.

The following proposition is a first step to understand Ω_C.

Proposition 4.1.13. $\dim_{k(C)} \Omega_C = 1$.

Like for functions in $k(C)^*$, one can associate to each $\omega \in \Omega_C$ a divisor as guaranteed by the following proposition.

Proposition 4.1.14. *Let $P \in C$ and let $t \in k(C)$ be such that* $\operatorname{ord}_P(t) = 1$.

(i) For each $\omega \in \Omega_C$, there exists a unique function $g \in k(C)$, depending on ω and t, such that $\omega = g\mathrm{d}t$. We denote g by $\omega/\mathrm{d}t$.

(ii) Assume $f \in k(C)$ is regular at P, i.e. $\operatorname{ord}_P(f) \geq 0$. *Then $\mathrm{d}f/\mathrm{d}t$ is also regular at P.*

(iii) *The quantity* $\mathrm{ord}_P(\omega/\mathrm{d}t)$ *depends only on* ω *and* P*; it is independent of the choice of* t*. Thus we can and will denote this quantity by* $\mathrm{ord}_P(\omega)$*.*

(iv) *Assume* $x, f \in k(C)$ *with* $x(P) = 0$*. Then* $\mathrm{ord}_P(f\mathrm{d}x) = \mathrm{ord}_P(f) + \mathrm{ord}_P(x) - 1$*.*

(v) *For all but finitely many* $P \in C$*, we have* $\mathrm{ord}_P(\omega) = 0$*.*

Thus we can make the following construction of divisors.

Definition 4.1.15. *Let* $\omega \in \Omega_C$*. The* ***divisor associated with*** ω *is*

$$\mathrm{div}(\omega) := \sum_{P \in C} \mathrm{ord}_P(\omega)[P] \in \mathrm{Div}(C).$$

Any such divisor, with $\omega \neq 0$*, is called a* ***canonical divisor****.*

The terminology "canonical divisor" is reasonable: Since $\dim_{k(C)} \Omega_C = 1$, each two canonical divisors are equivalent to each other.

Example 4.1.16. *Take the example from Example 4.1.7. Using* $\mathrm{d}x = \mathrm{d}(x-1) = \mathrm{d}(x+1)$ *and* $\mathrm{d}x = -x^2\mathrm{d}(1/x)$*, we get*

$$\mathrm{div}(\mathrm{d}x) = [P_1] + [P_2] + [P_3] - 3[\infty].$$

This is a canonical divisor of C*. Notice that* $\mathrm{div}(\mathrm{d}x/y) = 0$*. So* 0 *is also a canonical divisor for* C*.*

We finish this subsection by the following definition.

Definition 4.1.17. *A differential* $\omega \in \Omega_C$ *is said to be* ***regular*** *(or* ***holomorphic****) if* $\mathrm{div}(\omega) \geq 0$*, i.e.* $\mathrm{ord}_P(\omega) \geq 0$ *at all* $P \in C$*. By convention we say that* 0 *is a regular differential.*

Example 4.1.18. *Let* t *be a coordinate function on* $\mathbb{P}^1$*. It can be shown (Exercise class) that* $\mathrm{div}(\mathrm{d}t) = -2[\infty]$*. Thus* $\deg K_{\mathbb{P}^1} = -2 < 0$ *and hence* $\mathbb{P}^1$ *has no regular differentials.*

4.1.4 Genus and the Riemann–Roch Theorem

In this subsection, we assume C to be **smooth**.

Let $K_C \in \mathrm{Div}(C)$ be a canonical divisor on C, *i.e.* $K_C = \mathrm{div}(\omega)$ for some $\omega \in \Omega_C$. We have the following important invariant of C

Definition-Proposition 4.1.19. *The dimension* $\ell(K_C)$ *is independent of the choice of* K_C*.*
This dimension is called the ***genus*** *of the curve and is denoted by* $g(C)$ *(or simply* g*).*

Proof. To see that $\ell(K_C)$ is independent of the choice of K_C, it suffices to prove the following isomorphism of k-vector spaces:

$$L(K_C) \simeq \{\omega' \in \Omega_C : \omega' \text{ is regular}\}. \tag{4.1.3}$$

Let $0 \neq f \in L(K_C)$, then $\mathrm{div}(f) \geq -\mathrm{div}(\omega)$, and hence $\mathrm{div}(f\omega) \geq 0$. Therefore $f\omega \in \Omega_C$ is regular. We have thus established a map $L(K_C) \to \{\omega' \in \Omega_C : \omega' \text{ is regular}\} \cup \{0\}$, which is easily seen to be injective.

Let us prove the surjectivity. Since $\dim_{k(C)} \Omega_C = 1$, each differential ω' on C has the form $f\omega$ for some $f \in k(C)$. If ω' is regular, then $\mathrm{div}(f\omega) \geq 0$ and hence $f \in L(K_C)$. Hence ω' is the image of $f \in L(K_C)$ under the map defined above. Thus we have established the surjectivity. □

Example 4.1.20. *By Example 4.1.18, the curve* $\mathbb{P}^1$ *has genus* 0.

Example 4.1.21. *Take the example from Example 4.1.7 and Example 4.1.16. As* 0 *is a canonical divisor, we have* $g(C) = \ell(0) = 1$ *by Lemma 4.1.11.(ii).*

Theorem 4.1.22 (Riemann–Roch for curves)**.** *For each* $D \in \mathrm{Div}(C)$*, we have*

$$\ell(D) - \ell(K_C - D) = \deg D - g + 1.$$

As a corollary of the Riemann–Roch Theorem (applied to $D = K_C$) and Lemma 4.1.11.(ii), we have:

Corollary 4.1.23. $\deg K_C = 2g - 2$.

When the field $k = \mathbb{C}$, a more intuitive way to see the genus g is as follows. In this case, the smooth curve C is a Riemann surface, and g equals $\frac{1}{2}\mathrm{rank}_{\mathbb{Z}} H_1(C, \mathbb{Z})$.

4.1.5 A result of Weil

Let C be a curve of genus $g \geq 1$.

Lemma 4.1.24 (Weil)**.** *Let* $1 \leq d \leq g$ *be an integer. There exists a Zariski open dense subset* $U \subseteq C^d$ *such that* $\ell(\sum_{j=1}^d [P_j]) = 1$ *for all* $(P_1, \ldots, P_d) \in U$*. Equivalently,* $L(\sum_{j=1}^d [P_j]) = k$ *for all* $(P_1, \ldots, P_d) \in U$*.*

Proof. We start with the following preparation.
Claim 1: Let D be a divisor on C. Then $\ell(D - [P]) \geq \ell(D) - 1$ for all $P \in C$.

Indeed, it is easy to check that $L(D - [P]) \subseteq L(D)$. Assume $f_1, f_2 \in L(D) \setminus L(D - [P])$. Then $\mathrm{ord}_P(f_1) = \mathrm{ord}_P(f_2)$ which we assume to be m. Take a uniformizer t at P (*i.e.* a function $t \in k(C)^*$ such that $\mathrm{ord}_P(t) = 1$). Then locally at P we have $f_1 = a_1 t^m +$ higher terms and $f_2 = a_2 t^m +$ higher terms for some $a_1, a_2 \in k^*$. Thus $\mathrm{ord}_P(f_1 - (a_1/a_2)f_2) \geq m + 1$. By looking at all the points in $\mathrm{supp}(D)$, we then find that $f_1 - (a_1/a_2)f_2 \in L(D - [P])$. Thus we can conclude.
Claim 2: Let D be a divisor on C such that $\ell(D) \geq 1$. Then $\ell(D - [P]) = \ell(D) - 1$ for all P in a Zariski open dense subset of C.

Indeed, if we fix a function $0 \neq f \in L(D)$, then

$$\begin{aligned} \ell(D - [P]) = \ell(D) &\Rightarrow L(D - [P]) = L(D) \quad \text{since } L(D - [P]) \subseteq L(D) \\ &\Rightarrow f \in L(D - [P]) \\ &\Rightarrow \mathrm{div}(f) + D - [P] \geq 0 \\ &\Rightarrow P \in \mathrm{supp}(D) \text{ or } f(P) = 0. \end{aligned}$$

Thus $\{P \in C : \ell(D - [P]) = \ell(D)\}$ is contained in $\mathrm{supp}(D) \bigcup V(f)$ which is a finite set (since $f \neq 0$). So we can conclude by Claim 1.

Claim 3: Let D be a divisor on C. For each $d \in \{1, \ldots, g\}$, there exists a Zariski open dense subset U of C^d such that $\ell(D - \sum_{j=1}^d [P_j]) = \ell(D) - d$ for all $(P_1, \ldots, P_d) \in U$.

Indeed, let $U' = C \setminus \mathrm{supp}(D)$. Consider all points $(P_1, \ldots, P_d) \in (U')^d$ with the P_j's two-by-two distinct. We have the following exact sequence

$$0 \to L\left(D - \sum_{j=1}^d [P_j]\right) \to L(D) \to k^d$$

where the second map is the natural inclusion and the last map is the evaluation $f \mapsto (f(P_1), \ldots, f(P_d))$. Take a basis $\{f_1, \ldots, f_n\}$ of $L(D)$. We have $\ell(D - \sum_{j=1}^{d}[P_j]) = \ell(D) - d$ if and only if this evaluation is surjective, and hence if and only if $\det(f_i(P_j))_{i\in I,\ 1\leq .j\leq d} \neq 0$ for some $I \subseteq \{1, \ldots, n\}$ of cardinality d. This defines a Zariski open subset U of $(U')^d$, which is furthermore a Zariski open subset of C^d. Moreover U is non-empty by applying Claim 2 successively. Hence U is Zariski open dense, and we have established this claim.

Applying Claim 3 to the divisor K_C, we get a Zariski open dense subset U of C^d such that $\ell(K_C - \sum_{j=1}^{d}[P_j]) = \ell(K_C) - d = g - d$ for all $(P_1, \ldots, P_d) \in U$, and Riemann–Roch for curves (Theorem 4.1.22) implies $\ell(\sum_{j=1}^{d}[P_j]) = \ell(K_C - \sum_{j=1}^{d}[P_j]) + d - g + 1 = 1$ for all such $(P_1, \ldots, P_d)$. Now we are done. □

4.2 Curves and Jacobians

In this section, all varieties are defined over $\mathbb{C}$ unless otherwise stated.

By a *curve*, we mean an irreducible smooth projective curve.

Let C be a curve of genus g. It is a Riemann surface of genus g, *i.e.* $\mathrm{rank}_{\mathbb{Z}} H_1(C, \mathbb{Z}) = 2g$.

4.2.1 Periods

We will use $H^0(C, \Omega^1_C)$ to denote the set of holomorphic differentials; it is a $\mathbb{C}$-vector space of dimension g by (4.1.3).

Let γ be a path on the Riemann surface C and let $\omega \in H^0(C, \Omega^1_C)$, then one can compute the integral $\int_\gamma \omega$.

Example 4.2.1. *Take the example from Example 4.1.16 ($y^2 = x(x+1)(x-1) = x^3 - x$). We have seen that $g = g(C) = 1$, and thus $\omega := \mathrm{d}x/y$ is a basis of $H^1(C, \Omega^1_C)$ by the discussion in the example.*

Let γ be a path on C going from $(a, \sqrt{a^3 - a})$ to $(b, \sqrt{b^3 - b})$. Then the integral $\int_\gamma \omega$ on the Riemann surface C gives a precise meaning to the multivalued integral $\int_a^b 1/\sqrt{t^3 - t}\mathrm{d}t$; it is the choice of the path γ which has eliminated the indeterminacy.

A better way to understand the dependence of the integral on the path is via the homology. Take a basis $\{\gamma_1, \ldots, \gamma_g, \gamma_{g+1}, \ldots, \gamma_{2g}\}$ of $H_1(C, \mathbb{Z})$, which we assume to satisfy $\gamma_i \cdot \gamma_{j+g} = \delta_{ij}$ for each $1 \leq i, j \leq g$. If P and Q are two points in C, and γ and γ' are two paths joining P and Q, then $\gamma'^{-1} \circ \gamma$ is a closed path, and so is homologous to $\sum m_i\gamma_i$ for some integers m_i. Thus for any $\omega \in H^0(C, \Omega^1_C)$, we have

$$\int_\gamma \omega - \int_{\gamma'} \omega = \sum_{i=1}^{2g} m_i \int_{\gamma_i} \omega.$$

Now take a basis $\{\omega_1, \ldots, \omega_g\}$ of $H^1(C, \Omega^1_C)$. Then we have a $g \times 2g$ matrix

$$\Omega = (\Omega_1\ \Omega_2) = \left(\left(\int_{\gamma_i} \omega_j \right)_{1\leq i,j\leq g} \left(\int_{\gamma_{g+i}} \omega_j \right)_{1\leq i,j\leq g} \right). \tag{4.2.1}$$

We call Ω a *period matrix* of C, and let L_Ω be the $\mathbb{Z}$-module generated by the columns of Ω.

An important result is the following Riemann's period relations.

Theorem 4.2.2 (Riemann's period relations)**.** *We have*

$$\Omega_1\Omega_2^{\mathrm{t}} = \Omega_2\Omega_1^{\mathrm{t}} \quad \textit{and} \quad -\sqrt{-1}(\overline{\Omega}_1\Omega_2^{\mathrm{t}} - \overline{\Omega}_2\Omega_1^{\mathrm{t}}) > 0.$$

Here we write $M > 0$ for a $g \times g$ matrix to mean that M is positive definite (and hence symmetric).

A corollary of Riemann's period relations is that Ω_1 is invertible. Indeed, to see this, it suffices to prove: $\Omega_1^{\mathrm{t}} Y = 0 \Rightarrow Y = 0$.

Thus we can change the basis of $H^0(C, \Omega_C^1)$ to transform Ω_1 into the identity matrix I_g and Ω_2 into the matrix $\tau := \Omega_1^{-1}\Omega_2$. Thus we have a new period matrix $\Omega_{\mathrm{nor}} = (I_g \ \tau)$, and Riemann's relations say that τ is symmetric with positive definite imaginary part $\mathrm{Im}(\tau)$. Under this new basis, we have $L_{\Omega_{\mathrm{nor}}} = \mathbb{Z}^g + \tau\mathbb{Z}^g$. Notice that L_Ω is then $\Omega_1(\mathbb{Z}^g + \tau\mathbb{Z}^g)$. In particular, we have:

Corollary 4.2.3. *L_Ω is a lattice in $\mathbb{C}^g$.*

4.2.2 Jacobians

Definition 4.2.4. *The **Jacobian** of a curve C, denoted by $\mathrm{Jac}(C)$, is the complex torus $J(C) := \mathbb{C}^g/L_\Omega$, with L_Ω defined in the previous subsection.*

In fact, by the discussion at the end of the previous subsection, it suffices to take L_Ω to be $\mathbb{Z}^g + \tau\mathbb{Z}^g$, because the $g \times g$ invertible matrix $\Omega_1 \colon \mathbb{C}^g \to \mathbb{C}^g$ is an isomorphism of $\mathbb{C}^g$, which induces an isomorphism of complex tori $\mathbb{C}^g/(\mathbb{Z}^g + \tau\mathbb{Z}^g) \xrightarrow{\sim} \mathbb{C}^g/\Omega_1(\mathbb{Z}^g + \tau\mathbb{Z}^g)$. In particular, $J(C)$ is independent of the choice of the basis of $H^0(C, \Omega_C^1)$.

A more intrinsic formulation of the Jacobian is as follows. We can identify $H_1(C, \mathbb{Z})$ as a lattice in $H^0(C, \Omega_C^1)^\vee$, the dual of the $\mathbb{C}$-vector space $H^0(C, \Omega_C^1)$, via the map

$$H_1(C, \mathbb{Z}) \to H^0(C, \Omega_C^1)^\vee, \quad \gamma \mapsto \left(\gamma \mapsto \int_\gamma \omega\right).$$

Then the Jacobian of C is equal to

$$\mathrm{Jac}(C) = H^0(C, \Omega_C^1)^\vee / H_1(C, \mathbb{Z}). \tag{4.2.2}$$

For each $P \in C$, we can define a holomorphic map (called the ***Abel–Jacobi embedding via*** P)

$$j_P \colon C \to \mathrm{Jac}(C), \quad Q \mapsto \left(\int_P^Q \omega_1, \ldots, \int_P^Q \omega_g\right) \bmod L_\Omega. \tag{4.2.3}$$

The map j_P extends by linearity to $\mathrm{Div}(C) \to \mathrm{Jac}(C)$, $\sum n_Q[Q] \mapsto \sum n_Q j_P(Q)$. It should be understood that the sum on the right hand side is the sum on the complex torus (induced by the sum on $\mathbb{C}^g$), while the sum on the left hand side is just a formal sum. When restricted to $\mathrm{Div}^0(C)$, the set of divisors on C of degree 0, the map thus obtained

$$\Phi \colon \mathrm{Div}^0(C) \to \mathrm{Jac}(C)$$

is independent of the choice of P.

Theorem 4.2.5 (Abel–Jacobi)**.** *The map Φ defined above is a surjective group homomorphism, and $\mathrm{Ker}\Phi$ is precisely the subgroup of principal divisors. Thus $\mathrm{Cl}^0(C) \simeq \mathrm{Jac}(C)$, where $\mathrm{Cl}^0(C)$ is the divisor class group of divisors of degree 0.*

So far, we have seen that $\mathrm{Jac}(C)$ is a complex torus. An important result is:

Theorem 4.2.6. *$\mathrm{Jac}(C)$ is a projective variety, i.e. $\mathrm{Jac}(C)$ is an algebraic subvariety of $\mathbb{P}^N$ for some N. If $g(C) \geq 1$, then each j_P is a closed immersion.*

These theorems allow us to construct $\mathrm{Jac}(C)$ and the Abel–Jacobi map j_P in another way: We define $\mathrm{Jac}(C) := \mathrm{Cl}^0(C)$ and $j_P \colon C \to \mathrm{Jac}(C)$, $Q \mapsto \mathrm{cl}([Q]-[P])$. An advantage of this point of view is that these construction then generalize from over $k = \mathbb{C}$ to over an arbitrary field k of characteristic 0. For example if $k \subseteq \mathbb{C}$, then the group $\mathrm{Aut}(\mathbb{C}/k)$ acts on $\mathrm{Jac}(C_{\mathbb{C}}) = \mathrm{Cl}^0(C_{\mathbb{C}})$ and this endows $\mathrm{Cl}^0(C) = \mathrm{Jac}(C)$ with the structure of a projective variety defined over k. In particular, this applies to $k = \overline{\mathbb{Q}}$.

To prove that $\mathrm{Jac}(C)$ is a projective variety, one can use the knowledge on Riemann forms. Here we take a more algebro-geometric point of view. Let us temporarily assume that $g = g(C) \geq 2$. Then for each r, one can define the r-fold sum

$$W_r = W_r(C) := j_P(C) + \cdots + j_P(C) = \{j_P(Q_1) + \ldots + j_P(Q_r) : Q_1, \ldots, Q_r \in C\}. \quad (4.2.4)$$

It can be checked that $\dim W_r = \min\{r, g\}$. In particular, $\dim W_{g-1} = g-1$. This W_{g-1}, usually called a *Theta divisor* and denoted by Θ, gives rise to an embedding of $\mathrm{Jac}(C)$ into some $\mathbb{P}^N$. To understand this, we need to discuss about divisors on an arbitrary algebraic variety. This will be the content of the next section.

4.3 Weil and Cartier Divisors

The goal of this section is to gather some results for Weil and Cartier divisors on an arbitrary (quasi-projective) algebraic variety. Below, by k we always mean an algebraic closed field of characteristic 0. Typical examples are $k = \mathbb{C}$ or $k = \overline{\mathbb{Q}}$.

All algebraic varieties are assumed to be defined over k. We also make the following convention: algebraic varieties are assumed to be *irreducible*.

4.3.1 Weil divisors

Definition 4.3.1. *Let X be an algebraic variety. A **Weil divisor** on X is a finite formal sum of the form $D = \sum n_Y [Y]$, where $n_Y \in \mathbb{Z}$ and the Y's are subvarieties of X of codimension 1.*

*The **degree** of the Weil divisor D above is defined to be $\deg D := \sum_Y n_Y$.*

*The **support** of the Weil divisor D above, denoted by $\mathrm{supp}(D)$, is defined to be $\bigcup_{n_Y \neq 0} Y$.*

*A Weil divisor D as above is said to be **effective**, denoted by $D \geq 0$, if $n_Y \geq 0$ for all Y.*

For example when X is a projective curve, then the Y's are points; when X is a surface, then the Y's are irreducible curves.

When X satisfies some extra hypothesis (for example if X is non-singular or if X is normal), then one can define principal divisors (and hence the Weil divisor class group $\mathrm{Cl}(X)$) as for the case of smooth curves. We shall not go into details of this construction but simply say that they exist.

4.3.2 Cartier divisors

In general, it is often more convenient to work with another kind of divisors, called the *Cartier divisors*. Before giving the definition, let us see an example.

Example 4.3.2. *Consider the curve $\mathbb{P}^1$. It is covered by the Zariski open subsets $U_0 := \{[1:t] : t \in k\}$ and $U_1 := \{[t':1] : t' \in k\}$. We use $[x:y]$ to denote the homogeneous coordinates of $\mathbb{P}^1$.*

Consider the divisor $D = -[1:0] + [1:2] + [0:1]$.[2] When restricted to U_0, it becomes $D|_{U_0} = -[1:0] + [1:2]$ and is the divisor associated with the rational function $f_0 := (2x-y)/x$. When restricted to U_1, the divisor becomes $D|_{U_1} = [1:2] + [0:1]$ and hence is the divisor associated with the rational function $f_1 := y(2x-y)$.

Notice that on $U_0 \cap U_1 = \{[x:y] : xy \neq 0\}$, we have $f_0 f_1^{-1} = 1/xy$ has no zeros or poles on $U_0 \cap U_1$. Thus on $U_0 \cap U_1$, we have $\operatorname{div}(f_0 f_1^{-1})|_{U_0 \cap U_1} = 0$, and hence $\operatorname{div}(f_0)|_{U_0 \cap U_1} = \operatorname{div}(f_1)|_{U_0 \cap U_1}$.

In other words, the divisor D can be recovered by the data $\{(U_0, f_0), (U_1, f_1)\}$.

Definition 4.3.3. *A **Cartier divisor** on an algebraic variety X is an equivalence class of collections of pairs $\{(U_i, f_i) : i \in I\}$ satisfying the following conditions:*

(i) The U_i's are Zariski open subsets that cover X.

(ii) The f_i's are non-zero rational functions $f_i \in k(X)^*$.

(iii) For each $i, j \in I$, $f_i f_j^{-1}$ has no zeros or poles on $U_i \cap U_j$. In other words, $f_i f_j^{-1}$ is an invertible regular function on $U_i \cap U_j$.

Two collections $\{(U_i, f_i) : i \in I\}$ and $\{(V_j, g_j) : j \in J\}$ are said to be equivalent if $f_i g_j^{-1}$ has no zeros or poles on $U_i \cap V_j$ for each $i \in I$ and $j \in J$.

The sum of two Cartier divisors is defined by

$$\{(U_i, f_i) : i \in I\} + \{(V_j, g_j) : j \in J\} := \{(U_i \cap V_j, f_i g_j) : (i,j) \in I \times J\}.$$

With this operation, the Cartier divisors form a group which we denote by $\operatorname{CaDiv}(X)$. The ***support*** of a Cartier divisor D is the set of zeros of poles of the f_i's. A Cartier divisor D is said to be ***effective***, denoted by $D \geq 0$, if it can be defined by a collection $\{(U_i, f_i) : i \in I\}$ with each f_i having no poles on U_i.

Using the language of Cartier divisors, it is easy to define the principal divisors. Associated to each $f \in k(X)^*$, we can associate the Cartier divisor

$$\operatorname{div}(f) := \{(X, f)\}.$$

Definition 4.3.4. *A Cartier divisor $D \in \operatorname{CaDiv}(X)$ is **principal** if it equals $\operatorname{div}(f)$ for some $f \in k(X)^*$.*

*Two Cartier divisors D_1, D_2 are **linearly equivalent**, denoted $D_1 \sim D_2$, if $D_1 - D_2$ is a principal Cartier divisor.*

*The **Cartier divisor class group of** X, denoted by $\operatorname{CaCl}(X)$, is the group of Cartier divisors modulo linear equivalence.*

One can also define, for each Cartier divisor D, the k-vector space

$$L(D) := \{f \in k(X)^* : D + \operatorname{div}(f) \geq 0\} \cup \{0\}, \tag{4.3.1}$$

and check that $\ell(D) := \dim_k L(D)$ depends only on the Cartier divisor class.

As for Weil divisors on curves, it is not hard to check that $D \sim D' \Rightarrow \ell(D) = \ell(D')$.

[2] If we identify U_0 with $\mathbb{A}^1$ in the usual way and use ∞ to denote the point $[0:1]$, then $D = -[0] + [2] + [\infty]$.

Theorem 4.3.5. *There exist natural group homomorphisms*[3]

$$\mathrm{CaDiv}(X) \to \mathrm{Div}(X) \quad \textit{and} \quad \mathrm{CaCl}(X) \to \mathrm{Cl}(X).$$

Moreover, they are isomorphisms if X is non-singular.

Thus for smooth varieties, we will freely identify Weil and Cartier divisors in the rest of the course.

Cartier divisor class groups behave well under pullback.

Proposition 4.3.6. *Let $f\colon X \to Y$ be a morphism of varieties. Then there is a natural homomorphism $f^*\colon \mathrm{CaCl}(Y) \to \mathrm{CaCl}(X)$.*

Sketch. Let $\mathrm{Cl}(D) \in \mathrm{CaCl}(Y)$ be represented by $D \in \mathrm{CaDiv}(Y)$. The **Moving Lemma** says that there exists some $D' = \{(U_i, f_i) : i \in I\} \in \mathrm{CaDiv}(Y)$ such that $D' \sim D$ and $f(X) \not\subseteq \mathrm{supp}(D')$. We can then define a Cartier divisor

$$f^*D' := \{(f^{-1}(U_i), f_i \circ f) : i \in I\} \in \mathrm{CaDiv}(X).$$

Then we set $f^*\mathrm{Cl}(D)$ to be the class of f^*D' in $\mathrm{CaCl}(X)$.[4] □

4.3.3 Theta divisor on Jacobians

Let C be a curve of genus $g \geq 1$ and $P_0 \in C$. Use J to denote the Jacobian $\mathrm{Jac}(C)$, and let $j_{P_0}\colon C \to J$, $P \mapsto \mathrm{cl}([P] - [P_0])$ be the Abel–Jacobi embedding via P_0. For each $d \geq 1$, define the map

$$\Phi_d\colon C^d \to J, \quad (P_1, \ldots, P_d) \mapsto \mathrm{cl}\left(\sum_{i=1}^{d} [P_i] - d[P_0]\right) = \sum_{i=1}^{d} j_{P_0}(P_i).$$

Let Θ be the image of Φ_{g-1}. Then it has dimension $g-1$, and hence is a Weil divisor on J. Denote by $\Theta^- := [-1]^*\Theta$;[5] as a variety it is $-j_{P_0}(C) - \cdots - j_{P_0}(C)$ ($g-1$ copies). Both Θ and Θ^- are effective Weil divisors.

Since J is a smooth algebraic variety, by Theorem 4.3.5, one can identify $\mathrm{Div}(J) \simeq \mathrm{CaDiv}(J)$. Thus we will not distinguish Weil divisors and Cartier divisors on J. Similar on C.

Proposition 4.3.7. *There exists a Zariski open dense subset U of C^g satisfying the following property: $(P_1, \ldots, P_g) \in U \Rightarrow \sum_{i=1}^{g}[P_i] \sim j^*_{\Phi_g(P_1,\ldots,P_g)}(\Theta^-)$ as divisors on C, where $j_a\colon C \to J$ is defined by $P \mapsto j_{P_0}(P) - a$.*

Sketch of proof. By Lemma 4.1.24, there exists a Zariski open dense subset $U \subseteq C^g$ such that $L(\sum_{i=1}^{g}[P_i]) = k$ for all $(P_1, \ldots, P_g) \in U$. Shrink U such that each $(P_1, \ldots, P_g) \in U$ satisfies that $P_i \neq P_j$ for all $i \neq j$.

Let $a \in \Phi_g(U) \subset J \simeq \mathrm{Cl}^0(C)$. Let $D_a \in \mathrm{Div}^0(C)$ be such that $\mathrm{cl}(D_a) = a$. Then $\mathrm{cl}(D_a) = \mathrm{cl}(\sum_{i=1}^{g}[P_i] - g[P_0])$ for some $(P_1, \ldots, P_g) \in C^g$, and hence $g[P_0] + D_a \sim \sum_{i=1}^{g}[P_i]$.

Suppose $(P_1', \ldots, P_g') \in U$ also satisfies $g[P_0] + D_a \sim \sum_{i=1}^{g}[P_i]'$ (equivalently $\Phi_g(P_1', \ldots, P_g') = a$). Then as divisors on C we have $\sum_{i=1}^{g}[P_i] - \sum_{i=1}^{g}[P_i'] = \mathrm{div}(f)$ for some $f \in k(C)^*$. But

[3] Some extra conditions need to be put on X for the second map.

[4] For this construction to be well-defined, one uses the following fact: If $D, D' \in \mathrm{CaDiv}(Y)$ are linearly equivalent such that $f(X) \not\subseteq \mathrm{Supp}(D) \cup \mathrm{Supp}(D')$, then $f^*D \sim f^*D'$.

[5] Recall that over $\mathbb{C}$, we have $J = \mathbb{C}^g/L_\Omega$ for some lattice L_Ω in $\mathbb{C}^g$. The map $[-1]\colon J \to J$ is induced by the multiplication by -1 on $\mathbb{C}^g$.

then $f^{-1} \in L(\sum_{i=1}^g [P_i] - \sum_{i=1}^g [P_i']) \subseteq L(\sum_{i=1}^g [P_i]) = k$. So f is a non-zero constant and hence $\sum_{i=1}^g [P_i] = \sum_{i=1}^g [P_i']$.

Therefore we have the following conclusion: for all $a \in \Phi_g(U)$, $g[P_0]+D_a$ is linearly equivalent to exactly one effective divisor $\sum_{i=1}^g [P_i]$ on C; moreover the P_i's two by two distinct.

Now let us show that $a \in \Phi_g(U) \Rightarrow \sum_{i=1}^g [P_i] \sim j_a^*(\Theta^-)$ as divisors on C. Let $P \in \mathrm{supp}(j_a^*\Theta^-)$. We have $\mathrm{cl}([P] - [P_0]) - a = j_a(P) \in \Theta^- = -j_{P_0}(C) - \cdots - j_{P_0}(C)$. Let $D_a \in \mathrm{Div}^0(C)$ be such that $\mathrm{cl}(D_a) = a$. Then $[P]-[P_0]-D_a \sim -\sum_{i=1}^{g-1}[P_i'] + (g-1)[P_0]$ for some $(P_1', \ldots, P_{g-1}') \in C^{g-1}$, and so $[P] + \sum_{i=1}^{g-1}[P_i'] \sim g[P_0] + D_a$. Thus we have $[P] + \sum_{i=1}^{g-1}[P_i'] = \sum_{i=1}^g [P_i]$ (this is an equality!). So $P \in \{P_1, \ldots, P_g\}$. One also needs to check that each $[P_i]$ appears exactly once in $j_a^*\Theta$. We omit its proof but simply point out that it follows from an argument using the tangent spaces and the fact that the P_i's are two by two distinct. □

4.4 Line bundles and ampleness

The goal of this section is to gather some results for line bundles on an arbitrary (quasi-projective) algebraic variety. Below, by k we always mean an algebraic closed field of characteristic 0. Typical examples are $k = \mathbb{C}$ or $k = \overline{\mathbb{Q}}$.

All algebraic varieties are assumed to be defined over k. We also make the following convention: algebraic varieties are assumed to be *irreducible*.

4.4.1 Line bundles

Definition 4.4.1. *A **line bundle** on a variety X is an algebraic variety $\mathcal{L}$ endowed with a morphism $p\colon \mathcal{L} \to X$ with the following two properties:*

(i) Each fiber $\mathcal{L}_x = p^{-1}(x)$ is a k-vector space of dimension 1.

(ii) The fibration p is locally trivial. More precisely, for each $x \in X$, there exists an open neighborhood U of x such that there exists the following commutative diagram

$$\begin{array}{ccc} \mathcal{L}|_U := p^{-1}(U) & \xrightarrow[\sim]{\phi_U} & U \times \mathbb{A}^1 \\ {\scriptstyle p}\downarrow & \swarrow_{p_1} & \\ U & & \end{array}$$

The maps ϕ_U are called the local trivializations of $\mathcal{L}$.

An easy example is the *trivial line bundle* $X \times \mathbb{A}^1 \to X$ on X.

Example 4.4.2. *Here is an example of non-trivial line bundle over $\mathbb{P}^n$. View $\mathbb{P}^n$ as the set of lines in $\mathbb{A}^{n+1}$ passing through 0. Define*

$$\mathcal{O}(-1) := \{(x, v) \in \mathbb{P}^n \times \mathbb{A}^{n+1} : v \text{ lies in the line } x\}.$$

Then the projection onto the first factor $p\colon \mathcal{O}(-1) \to \mathbb{P}^n$ gives $\mathcal{O}(-1)$ the structure of a line bundle. Indeed, condition (i) is clear. For (ii), the fibration p can be trivialized over each standard affine open chart $U_j := \mathbb{P}^n \setminus V(X_j)$, with the trivialization given by

$$U_j \times \mathbb{A}^1 \to \mathcal{L}|_{U_j}, \quad (x, \lambda) \mapsto \left(x, \left(\frac{\lambda x_0}{x_j}, \ldots, \frac{\lambda x_n}{x_j}\right)\right).$$

Definition 4.4.3. *A **morphism** between two line bundles $p\colon \mathcal{L} \to X$ and $p'\colon \mathcal{L}' \to X$ on X is a morphism $\Phi\colon \mathcal{L} \to \mathcal{L}'$ such that $p = p' \circ \Phi$ and that the map $\Phi_x\colon \mathcal{L}_x \to \mathcal{L}'_x$ is a linear transformation of $\mathbb{A}^1$ for each $x \in X$.*

With this definition, we can define ***isomorphism classes line bundles on*** X, the set of which is denoted by $\mathrm{Pic}(X)$.

Next we describe how to construct line bundles by gluing locally trivial line bundles. The basic observation is that a line bundle $\mathcal{L}$ and local trivializations ϕ_U fit into the following commutative diagram

$$\begin{array}{ccccc} \mathcal{L}|_{U_i\cap U_j} & \xrightarrow[\sim]{\phi_{U_i}} & (U_i\cap U_j)\times\mathbb{A}^1 & \xleftarrow[\sim]{\phi_{U_j}} & \mathcal{L}|_{U_i\cap U_j} \\ & \searrow^{p} & \downarrow & \swarrow_{p} & \\ & & U_i\cap U_j & & \end{array}$$

We thus obtain isomorphisms $\phi_{U_j} \circ \phi_{U_i}^{-1}\colon (U_i \cap U_j) \times \mathbb{A}^1 \to (U_i \cap U_j) \times \mathbb{A}^1$ that must be of the shape $(x, v) \mapsto (x, g_{ji}(x)(v))$, with g_{ji} being a regular function on $U_i \cap U_j$. These g_{ji}'s are called the *transition functions*. The following identities are immediate:

$$g_{ii} = \mathrm{id} \quad \text{and } g_{ij}g_{jk} = g_{ik} \text{ on } U_i \cap U_j \cap U_k.$$

The set of g_{ji}'s determines the line bundle $\mathcal{L}$. Conversely, any set of g_{ji}'s satisfying these identities can be used to construct a line bundle by gluing together (local) trivial line bundles.

From this construction, one can define:

(i) The ***dual*** of a line bundle $\mathcal{L}$ to be $\mathcal{L}^{\otimes -1}$ whose fibers are the dual vector spaces.

(ii) The ***tensor product*** of two line bundle $\mathcal{L}$ and $\mathcal{L}'$ to be $\mathcal{L} \otimes \mathcal{L}'$ whose fibers are the tensor products of the fibers of $\mathcal{L}$ and of $\mathcal{L}'$.

(iii) The ***pullback*** of a line bundle $p\colon \mathcal{L} \to X$ by a morphism $f\colon Y \to X$ to be the fiber product

$$f^*\mathcal{L} := Y \times_X \mathcal{L} = \{(y, v) \in Y \times \mathcal{L} : f(y) = p(v)\}.$$

In particular, $f^*(X \times \mathbb{A}^1) = Y \times \mathbb{A}^1$.

Remark 4.4.4. *The tensor product and the dual endow* $\mathrm{Pic}(X)$ *with the structure of abelian groups. So we call* $\mathrm{Pic}(X)$ *the **group of isomorphism classes of line bundles on** X.*

Definition 4.4.5. *Let $p\colon \mathcal{L} \to X$ be a line bundle. A **section** of $\mathcal{L}$ is a morphism $s\colon X \to \mathcal{L}$ such that $p \circ s = \mathrm{id}_X$. Similarly a **rational section** of $\mathcal{L}$ is a rational map $s\colon X \dashrightarrow \mathcal{L}$ such that $p \circ s = \mathrm{id}_X$.*

The set of sections of a line bundle $\mathcal{L}$ will be denoted by $H^0(X, \mathcal{L})$; it is a k-vector space of finite dimension.

Remark 4.4.6. *A more concrete way to understand sections s of $\mathcal{L}$ is as follows. Write $\phi_{U_i}\colon \mathcal{L}|_{U_i} \simeq U_i \times \mathbb{A}^1$ for the local trivializations with the transition functions g_{ji}'s. Then s can be identified with a collection $\{s_i\colon U_i \to \mathbb{A}^1$ with $g_{ji}s_i = s_j$ on $U_i \cap U_j\}_{i\in I}$. Locally on each U_i, $\phi_{U_i} \circ s|_{U_i}\colon U_i \to U_i \times \mathbb{A}^1$ is then $x \mapsto (x, s_i(x))$.*

For example, $H^0(X, X \times \mathbb{A}^1) = \mathbb{A}^1$ for any projective variety X. For Example 4.4.2, one can check $H^0(\mathbb{P}^n, \mathcal{O}(-1)) = 0$.

For any morphism $f\colon Y \to X$ and any line bundle $\mathcal{L}$ on X, there is a natural morphism

$$H^0(X, \mathcal{L}) \to H^0(Y, f^*\mathcal{L}), \quad s \mapsto (y \mapsto (y, s \circ f(y))). \tag{4.4.1}$$

4.4.2 Line bundles and Cartier divisors

To each Cartier divisor $D = \{(U_i, f_i) : i \in I\}$, one can associate a line bundle $\mathcal{O}(D) \to X$ by gluing the trivial line bundle $U_i \times \mathbb{A}^1 \to U_i$ via the transition functions $g_{ji} = f_j f_i^{-1}$

$$(U_i \cap U_j) \times \mathbb{A}^1 \to (U_i \cap U_j) \times \mathbb{A}^1, \quad (x, \lambda) \mapsto (x, \lambda \cdot (f_j f_i^{-1})(x)).$$

Since replacing the f_i's by $f_i f$ does not affect this construction, we obtain a homomorphism $\mathrm{CaCl}(X) \to \mathrm{Pic}(X)$.

Theorem 4.4.7. *The homomorphism* $\mathrm{CaCl}(X) \to \mathrm{Pic}(X)$ *induced by* $D \mapsto \mathcal{O}(D)$ *above is an isomorphism. More precisely,* $\mathcal{O}(D + D') = \mathcal{O}(D) \otimes \mathcal{O}(D')$ *and* $\mathcal{O}(-D) = \mathcal{O}(D)^\vee$.

Moreover, $\mathcal{O}(f^*D) = f^*\mathcal{O}(D)$ *for any morphism* $f\colon Y \to X$ *of varieties.*

To recover the Cartier divisor from a line bundle $\mathcal{L}$, one takes a rational section s of $\mathcal{L}$ and set $D = \mathrm{div}(s)$.

Proposition 4.4.8. *With the setting of the language in Theorem 4.4.7, there exists a natural bijection between* $H^0(X, \mathcal{O}(D))$ *and* $L(D)$.

Sketch. Write $D = \{(U_i, f_i) : i \in I\}$ for the Cartier divisor.

Take $0 \neq f \in L(D)$. Let us explain how to construct to the section $s_f \in H^0(X, \mathcal{O}(D))$ associated with f. The local trivializations of $\mathcal{O}(D)$ are $\phi_{U_i}\colon \mathcal{O}(D)|_{U_i} \simeq U_i \times \mathbb{A}^1$. Let $s_i\colon U_i \to \mathcal{L}_{U_i}$ be the composite of $U_i \to U_i \times \mathbb{A}^1$, $x \mapsto (x, (ff_i)(x))$, and $\phi_{U_i}^{-1}$. Since $D + \mathrm{div}(f) \geq 0$ by definition of $L(D)$, the product ff_i has no poles on U_i. By Remark 4.4.6, these s_i's patch together into $s_f \in H^0(X, \mathcal{O}(D))$.

Conversely for $0 \neq s \in H^0(X, \mathcal{O}(D))$, the rational function $f_s \in k(X)^*$ is constructed as follows. By Remark 4.4.6 s corresponds to $\{s_i\colon U_i \to \mathbb{A}^1\}_{i \in I}$ such that $f_j f_i^{-1} s_i = s_j$. Thus the functions $f_i^{-1} s_i$'s patch together into a function in $k(X)^*$. This is the desired f_s. □

Example 4.4.9. *Let* H *be a hyperplane in* $\mathbb{P}^n$*, and use* $\mathcal{O}(1)$ *to denote the associated line bundle. Then* $\mathcal{O}(1)$ *is the dual of* $\mathcal{O}(-1)$ *defined in Example 4.4.2. We have*

$$H^0(\mathbb{P}^n, \mathcal{O}(1)) = kX_0 \oplus \cdots \oplus kX_n = \{\textit{homogeneous polynomials of degree } 1\}.$$

For each $d \in \mathbb{N}$*, let* $\mathcal{O}(d) := \mathcal{O}(1) \otimes \cdots \otimes \mathcal{O}(1)$ *(d-times). Then*

$$H^0(\mathbb{P}^n, \mathcal{O}(d)) = \{\textit{homogeneous polynomials of degree } d\}.$$

4.4.3 Polynomials viewed as sections of line bundles

In Example 4.4.9, we saw that homogeneous polynomials of degree d in $n+1$ variables are precisely the sections of the line bundle $\mathcal{O}(d)$ on $\mathbb{P}^n$. This is an important point of view to understand polynomials and it deserves further discussion.

Now let us consider a bi-homogeneous polynomial $P \in k[X_0, \ldots, X_n; Y_0, \ldots, Y_m]$ of bi-degree (d_1, d_2). Consider the following line bundle on $\mathbb{P}^n \times \mathbb{P}^m$. Let $\pi_1\colon \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^n$ and $\pi_2\colon \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^m$ be the natural projections. Then $\pi_1^*\mathcal{O}(d_1) \otimes \pi_2^*\mathcal{O}(d_2)$ is a line bundle on $\mathbb{P}^n \times \mathbb{P}^m$, which we denote by $\mathcal{O}(d_1, d_2)$. Then $H^0(\mathbb{P}^n \times \mathbb{P}^m, \mathcal{O}(d_1, d_2))$ can be identified with the set of bi-homogeneous polynomials in $(n+1, m+1)$ variables of bi-degree (d_1, d_2).

This discussion can be generalized to an arbitrary product of projective spaces. Let

$$\mathbb{P} := \mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_r}.$$

Let $\pi_j\colon \mathbb{P} \to \mathbb{P}^{n_j}$ be the natural projection to the j-th factor. For any $d_1, \dots, d_r$ positive integers, we define the line bundle $\mathcal{O}(d_1, \dots, d_r)$ on $\mathbb{P}$ to be $\bigotimes_{j=1}^r \pi_j^*\mathcal{O}(d_j)$. Then $H^0(\mathbb{P}, \mathcal{O}(d_1, \dots, d_r))$ can be identified with the set of multi-homogeneous polynomials in $(n_1+1, \dots, n_r+1)$ variables of multi-degree $(d_1, \dots, d_r)$.

This can be applied to an arbitrary polynomial $f \in k[X_1, \dots, X_r]$ of partial degrees $d_1, \dots, d_r$ in the following way. Notice that f is a regular function on the affine variety k^r. We can embed k^r into $(\mathbb{P}^1)^r$ using the natural embedding $k \subseteq \mathbb{P}^1$, $t \mapsto [t:1]$. Now, if we use $[Y_0^{(j)} : Y_1^{(j)}]$ to denote the coordinates of the j-th $\mathbb{P}^1$, then f gives rise to a multi-homogeneous polynomial $F \in k[Y_0^{(1)}, Y_1^{(1)}; \dots; Y_0^{(r)}, Y_1^{(r)}]$ of multi-degree $(d_1, \dots, d_r)$, and we have seen that F is a section of the line bundle $\mathcal{O}(d_1, \dots, d_r)$ on $(\mathbb{P}^1)^r$. To recover f from F, it suffices to set all the $Y_1^{(j)}$'s to be 1. In other words, every polynomial can be seen as a section of an appropriate line bundle on a product of projective spaces.

Another application of the discussion above is the following lemma. We leave the proof as an exercise.

Lemma 4.4.10. *For the Segre embedding* (1.2.5) $S_{n,m}\colon \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{(n+1)(m+1)-1}$, $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} \otimes \mathbf{y} := (x_i y_j)_{i,j}$, *we have* $S_{n,m}^*\mathcal{O}(1) \simeq \mathcal{O}(1,1)$.

4.4.4 Ampleness

Let $\mathcal{L}$ be a line bundle on X. Take a basis $\{s_0, \dots, s_n\}$ of the k-vector space $H^0(X, \mathcal{L})$ (of dimension $n+1$). Then we obtain a rational map

$$\phi_{\mathcal{L}}\colon X \dashrightarrow \mathbb{P}^n, \qquad x \mapsto [s_0(x) : \cdots : s_n(x)].$$

Definition 4.4.11. *The line bundle $\mathcal{L}$ is said to be* ***very ample*** *if $\phi_{\mathcal{L}}$ is an immersion with $\mathcal{L} \simeq \phi_{\mathcal{L}}^*\mathcal{O}(1)$, and is said to be* ***ample*** *if $\mathcal{L}^{\otimes N}$ is very ample for some $N \geq 1$.*

Notice that the map $\phi_{\mathcal{L}}$ depends on the choice of the basis, but the property of $\mathcal{L}$ being (very) ample or not is independent of such choices.

We have a similar definition for divisors. Let D be a Cartier divisor. It is said to be ***(very) ample*** if and only if $\mathcal{O}(D)$ is. We also use ϕ_D to denote $\phi_{\mathcal{O}(D)}$.

Example 4.4.12. *For example, $\mathcal{O}(1)$ is very ample on $\mathbb{P}^n$ and an immersion can be obtained by taking $\phi_{\mathcal{O}(1)}\colon \mathbb{P}^n \to \mathbb{P}^n$ to be the identity map, and $\mathcal{O}(d)$ is very ample on $\mathbb{P}^n$ for all $d \geq 1$ with the immersion $\phi_{\mathcal{O}(d)}\colon \mathbb{P}^n \to \mathbb{P}^N$ being the d-uple embedding.*

Proposition 4.4.13. *Let $\iota\colon Y \to X$ be a finite morphism (for example, an immersion) and let $\mathcal{L} \in \mathrm{Pic}(X)$ be an ample line bundle. Then $\iota^*\mathcal{L}$ is ample on Y.*

Proposition 4.4.14. *Let $\mathcal{L}$ and $\mathcal{M}$ be two line bundles on X. Assume that $\mathcal{M}$ is ample. Then there exists $N \gg 1$ such that $\mathcal{L} \otimes \mathcal{M}^{\otimes N}$ is ample.*

Corollary 4.4.15. *Each line bundle $\mathcal{L}$ on a quasi-projective variety X can be written as $\mathcal{L}_1 \otimes \mathcal{L}_2^\vee$ for some ample line bundles $\mathcal{L}_1$ and $\mathcal{L}_2$ on X.*

Here is an important ampleness result concerning Jacobians and Theta divisors. Let C be a curve of genus $g \geq 1$ and $P_0 \in C$. Use J to denote the Jacobian $\mathrm{Jac}(C)$, and let $j_{P_0}\colon C \to J$, $P \mapsto \mathrm{cl}([P] - [P_0])$ be the Abel–Jacobi embedding via P_0. Let $\Theta := j_{P_0}(C) + \cdots + j_{P_0}(C)$ ($g-1$ copies). Then it has dimension $g-1$, and hence is a Weil divisor on J. Denote by $\Theta^- := [-1]^*\Theta$; as a variety it is $-j_{P_0}(C) - \cdots - j_{P_0}(C)$ ($g-1$ copies). Both Θ and Θ^- are effective Weil divisors. Since J is a smooth algebraic variety, by Theorem 4.3.5 Θ and Θ^- are also Cartier divisors.

Theorem 4.4.16. *Both Θ and Θ^- are ample on J.*

Here only the ampleness of Θ needs to be proved; then one uses Proposition 4.4.13 (applied to $[-1]\colon J \to J$) to deduce the ampleness of Θ^-.

4.5 Abelian varieties

In this section, we gather some properties for the general abstract theory of abelian varieties. An important example is the Jacobians of curves.

Let k be an algebraically closed field of characteristic 0. Typical examples are $k = \mathbb{C}$ and $k = \overline{\mathbb{Q}}$.

All algebraic varieties are assumed to be defined over k. We also make the following convention: algebraic varieties are assumed to be *irreducible*.

4.5.1 Abstract definition and abelian varieties over $\mathbb{C}$

Let G be an algebraic variety.

Definition 4.5.1. *The variety G is called an **algebraic group** (over k) if there exist*

- *a morphism $m\colon G \times G \to G$ (multiplication);*
- *a morphism $\iota\colon G \to G$ (inverse);*
- *a point $\epsilon \in G$ (identity);*

such that G is a group with multiplication, inverse, identity induced by m, ι, ϵ.

Example 4.5.2. *Let $J = \mathbb{C}^g/L_\Omega$ be a Jacobian defined over $\mathbb{C}$. It is known that J is an algebraic variety. Moreover, J has a natural structure of groups induced by the addition on $\mathbb{C}^g$, the negation on $\mathbb{C}^g$, and the origin $0 \in \mathbb{C}^g$. Thus J is an algebraic group (over $\mathbb{C}$).*

For arbitrary k, if $J = \mathrm{Cl}^0(C)$ is a Jacobian defined over k, then J is also an algebraic group over k.

Definition 4.5.3. *An **abelian variety** is a projective irreducible algebraic group.*

This definition is too abstract, although the requirements are easy to say. By this definition and the example above, each Jacobian is an abelian variety.

By knowledge of Lie groups, one can show:

Proposition 4.5.4. *Any abelian variety over $\mathbb{C}$ is a complex torus, i.e. equals $\mathbb{C}^g/\Lambda$ for some lattice $\Lambda \subseteq \mathbb{C}^g$.*

The converse of this proposition is not true unless $g = 1$, that is, when $g \geq 2$, then there exist lattices Λ such that $\mathbb{C}^g/\Lambda$ does not admit a complex-analytic embedding into some projective space $\mathbb{P}^N(\mathbb{C})$ (for any N). In fact, we have the following theorem.

Theorem 4.5.5. *Let Λ be a lattice in $\mathbb{C}^g$. Then $\mathbb{C}^g/\Lambda$ is an abelian variety if and only if there exists a positive definite Hermitian form $H\colon \mathbb{C}^g \times \mathbb{C}^g \to \mathbb{C}$ such that its imaginary part satisfies $\mathrm{Im}(H)(\Lambda \times \Lambda) \subseteq \mathbb{Z}$.*

Let $A = \mathbb{C}^g/\Lambda$ be an abelian variety over $\mathbb{C}$. For each $n \in \mathbb{Z}$, the multiplication $n\cdot\colon \mathbb{C}^g \to \mathbb{C}^g$, $x \mapsto nx$, induces a morphism $A \to A$ which we denote by

$$[n]\colon A \to A. \tag{4.5.1}$$

It is not hard to check that $[n]$ is a group homomorphism, which is surjective and has kernel isomorphic to $\mathbb{Z}^{2g}/n\mathbb{Z}^{2g}$.

For a general k, one possible way to understand an abelian variety A over k is as follows. Let us stick to the case $k = \overline{\mathbb{Q}} \subseteq \mathbb{C}$. By definition, $A \subseteq \mathbb{P}^N(\overline{\mathbb{Q}})$ for some $N \geq 1$. Then there exist homogeneous polynomials $f_1, \ldots, f_m$ with coefficients in $\overline{\mathbb{Q}}$ such that A is the zero locus $V(f_1, \ldots, f_m)$. Denote by $A(\mathbb{C})$ the $\mathbb{C}$-solutions to the system $f_1 = \cdots = f_m = 0$. Then $A(\mathbb{C})$ is an abelian variety over $\mathbb{C}$, and then we using the action of $\mathrm{Aut}(\mathbb{C}/\overline{\mathbb{Q}})$ to go back to A.

For example, let A be an abelian variety defined over $\overline{\mathbb{Q}}$. For each $n \in \mathbb{Z}$, we have defined above a morphism $[n]\colon A(\mathbb{C}) \to A(\mathbb{C})$. One can check that this morphism is invariant under the action of $\mathrm{Aut}(\mathbb{C}/\overline{\mathbb{Q}})$ on $A(\mathbb{C})$. Thus we obtain a morphism

$$[n]\colon A \to A, \tag{4.5.2}$$

which is again a group homomorphism. It is still surjective whose kernel is isomorphic to $\mathbb{Z}^{2g}/n\mathbb{Z}^{2g}$.

Each abelian variety A is smooth. Hence we will not distinguish Weil divisors and Cartier divisors on A by Theorem 4.3.5.

4.5.2 Theorem of the cube

Let A be an abelian variety.

For each subset $I \subseteq \{1, 2, 3\}$, set

$$s_I\colon A \times A \times A \to A, \quad (x_1, x_2, x_3) \mapsto \sum_{i \in I} x_i.$$

For example, if $I = \{1\}$, then s_1 is the projection to the first factor; if $I = \{1, 2\}$, $s_{12}(x_1, x_2, x_3) = x_1 + x_2$; if $I = \{1, 2, 3\}$, then $s_{123} = x_1 + x_2 + x_3$.

Theorem 4.5.6. *Let $L \in \mathrm{Pic}(A)$ be a line bundle on A. Then $s_{123}^*L \otimes s_{12}^*L^{\otimes -1} \otimes s_{13}^*L^{\otimes -1} \otimes s_{23}^*L^{\otimes -1} \otimes s_1^*L \otimes s_2^*L \otimes s_3^*L$ is isomorphic to the trivial line bundle on $A \times A \times A$.*

In the language of divisors (see Theorem 4.4.7 for the translation), the theorem is equivalent to: Let D be a divisor on A. Then $s_{123}^*D - s_{12}^*D - s_{13}^*D - s_{23}^*D + s_1^*D + s_2^*D + s_3^*D \sim 0$ as divisors on $A \times A \times A$.

Corollary 4.5.7. *Let V be an arbitrary variety, and let $f, g, h\colon V \to A$ be three morphisms. Then for any $L \in \mathrm{Pic}(A)$, the line bundle $(f+g+h)^*L \otimes (f+g)^*L^{\otimes -1} \otimes (f+h)^*L^{\otimes -1} \otimes (g+h)^*L^{\otimes -1} \otimes f^*L \otimes g^*L \otimes h^*L$ is isomorphic to the trivial line bundle on V.*

Proof. Denote by $\mathrm{cube}(L) := s_{123}^*L \otimes s_{12}^*L^{\otimes -1} \otimes s_{13}^*L^{\otimes -1} \otimes s_{23}^*L^{\otimes -1} \otimes s_1^*L \otimes s_2^*L \otimes s_3^*L$ as in Theorem 4.5.6. Then for the morphism $(f, g, h)\colon V \to A \times A \times A$, we have

$$(f+g+h)^*L \otimes (f+g)^*L^{\otimes -1} \otimes (f+h)^*L^{\otimes -1} \otimes (g+h)^*L^{\otimes -1} \otimes f^*L \otimes g^*L \otimes h^*L \simeq (f, g, h)^*\mathrm{cube}(L).$$

But $\mathrm{cube}(L)$ is isomorphic to the trivial line bundle $(A \times A \times A) \times \mathbb{A}^1$ by Theorem 4.5.6. Hence we are done. □

Corollary 4.5.8. *Let $L \in \mathrm{Pic}(A)$, and denote by $L_- := [-1]^*L$. Then we have*

$$[n]^*L \simeq L^{\otimes \frac{n^2+n}{2}} \otimes L_-^{\otimes \frac{n^2-n}{2}}, \quad \text{for all } n \in \mathbb{Z}.$$

As usual, we set $L^{\otimes 0}$ to be the trivial line bundle.

Proof. Apply Corollary 4.5.7 to $V = A$, $f = [n]$, $g = [1]$, and $h = [-1]$. Then we have

$$[n+1]^*L \otimes [n-1]^*L \otimes [n]^*L^{\otimes -2} \simeq L \otimes [-1]^*L.$$

Now the conclusion follows from an induction, both upwards and downwards, from $n = 0$. □

4.5.3 Theorem of square

Let A be an abelian variety. For each $a \in A$, set

$$t_a \colon A \to A, \quad x \mapsto a + x.$$

Then t_a is a morphism of algebraic varieties. It is called the *translation by a*.

Theorem 4.5.9 (Theorem of the square)**.** *For all $D \in \mathrm{Div}(A)$, $a, b \in A$, we have*

$$t_{a+b}^*D + D \sim t_a^*D + t_b^*D$$

as divisors on A.

Proof. Apply Corollary 4.5.7 to $L = \mathcal{O}(D)$, $V = A$, $f(x) = x$, $g(x) = a$ and $h(x) = b$. Then we have that $t_{a+b}^*\mathcal{O}(D) \otimes \mathcal{O}(D) \simeq t_a^*\mathcal{O}(D) \otimes t_b^*\mathcal{O}(D)$. Hence we are done. □

An application of the theorem of the square is the following result regarding Theta divisors on Jacobians.

Let C be a curve of genus $g \geq 1$ and $P_0 \in C$. Use J to denote the Jacobian $\mathrm{Jac}(C)$, and let $j_{P_0} \colon C \to J$, $P \mapsto \mathrm{cl}([P] - [P_0])$ be the Abel–Jacobi embedding via P_0. For each $d \geq 1$, define the map

$$\Phi_d \colon C^d \to J, \quad (P_1, \ldots, P_d) \mapsto \mathrm{cl}\left(\sum_{i=1}^{d}[P_i] - d[P_0]\right) = \sum_{i=1}^{d} j_{P_0}(P_i). \tag{4.5.3}$$

Let Θ be the image of Φ_{g-1}. Then it has dimension $g - 1$, and hence is a Weil divisor on J. Denote by $\Theta^- := [-1]^*\Theta$; as a variety it is $-j_{P_0}(C) - \cdots - j_{P_0}(C)$ ($g - 1$ copies). By Theorem 4.4.16 Θ and Θ^- are ample divisors on J.

Proposition 4.5.10. *For all $(P_1, \ldots, P_g) \in C^g$, we have the equivalence of divisors on C*

$$\sum_{i=1}^{g}[P_i] \sim j^*_{\Phi_g(P_1,\ldots,P_g)}(\Theta^-)$$

with $j_{\Phi_g(P_1,\ldots,P_g)} \colon C \to J$, $P \mapsto j_{P_0}(P) - \Phi_g(P_1, \ldots, P_g)$.

Proof. By Proposition 4.3.7, there exists a Zariski open dense subset U of C^g satisfying the following property: $(P_1, \ldots, P_g) \Rightarrow \sum_{i=1}^{g}[P_i] \sim j^*_{\Phi_g(P_1,\ldots,P_g)}(\Theta^-)$ as divisors on C, where $j_a \colon C \to J$ is defined by $P \mapsto j_{P_0}(P) - a$.

It can be shown using general algebraic geometry knowledge that $\Phi_g(U)$ contains a Zariski open dense subset V of J (by looking at dimensions, for example). Hence we are done.

Now we prove the desired conclusion. First, notice that $V + V - V = J$, *i.e.* the morphism $V \times V \times V \to J$, $(a, b, c) \mapsto a + b - c$, is surjective. Indeed, the map $(b, c) \mapsto b - c$ is already surjective because for all $x \in J$, we have $(V - x) \cap V \neq$ and hence $x = v - u$ for some $u, v \in V$.

Next for each $(P_1, \dots, P_g) \in C^g$, write $a = \Phi_g(P_1, \dots, P_g)$. Then $a = a_1 + a_2 - a_3$ with $a_l \in V \subseteq \Phi_g(U)$. Then by Theorem 4.5.9, we have[6]

$$t_{-a}^*\Theta^- = t_{-a_1-a_2+a_3}^*\Theta^- \sim t_{-a_1}^*\Theta^- + t_{-a_2}^*\Theta^- - t_{-a_3}^*\Theta^-.$$

Since $j_a = t_{-a} \circ j_{P_0}$ for all $a \in J$, applying $j_{P_0}^*$ to the linear equivalence above we obtain

$$j_a^*\Theta^- \sim j_{a_1}^*\Theta^- + j_{a_2}^*\Theta^- - j_{a_3}^*\Theta^-.$$

Since $a_l \in \Phi_g(U)$, we can write $a_l = \Phi_g(P_1^{(l)}, \dots, P_g^{(l)})$ for $l \in \{1, 2, 3\}$. Then the conclusion for the points in $\Phi_g(U)$ yields $\sum_{i=1}^g [P_i^{(l)}] \sim j_{a_l}^*(\Theta^-)$ for each $l \in \{1, 2, 3\}$. Hence

$$j_a^*\Theta^- \sim \sum_{i=1}^g [P_i^{(1)}] + \sum_{i=1}^g [P_i^{(2)}] - \sum_{i=1}^g [P_i^{(3)}]. \tag{4.5.4}$$

But $\Phi_g(P_1, \dots, P_g) = a = a_1 + a_2 - a_3 = \Phi_g(P_1^{(1)}, \dots, P_g^{(1)}) + \Phi_g(P_1^{(2)}, \dots, P_g^{(2)}) - \Phi_g(P_1^{(3)}, \dots, P_g^{(3)})$. In view of the definition of Φ_g, this means

$$\sum_{i=1}^g [P_i] - g[P_0] \sim \left(\sum_{i=1}^g [P_i^{(1)}] - g[P_0]\right) + \left(\sum_{i=1}^g [P_i^{(2)}] - g[P_0]\right) - \left(\sum_{i=1}^g [P_i^{(3)}] - g[P_0]\right).$$

So $\sum_{i=1}^g [P_i] \sim \sum_{i=1}^g [P_i^{(1)}] + \sum_{i=1}^g [P_i^{(2)}] - \sum_{i=1}^g [P_i^{(3)}]$. Combined with (4.5.4), we are done. □

4.5.4 Poincaré divisor class

Let C be a curve of genus $g \geq 1$. Let $P_0 \in C$. Let $j_{P_0} \colon C \to J$ be the Abel–Jacobi embedding via P_0. Let Θ be the theta divisor on J defined under (4.5.3).

Let $\Delta \subseteq C \times C$ be the diagonal. Then Δ is a Weil divisor on $C \times C$. By Theorem 4.3.5 it can be viewed as a Cartier divisor.

We define three morphisms $J \times J \to J$: $m(x, y) = x + y$, $p_1(x, y) = x$, and $p_2(x, y) = y$.

Set $\boldsymbol{\delta} := m^*\Theta - p_1^*\Theta - p_2^*\Theta$, which is a divisor on $J \times J$.

The goal of this subsection is to prove the following theorem. It will play an important role in the proof of Faltings's Theorem at the end of this course.

Theorem 4.5.11. *As divisors on $C \times C$, we have $(j_{P_0} \times j_{P_0})^*\boldsymbol{\delta} \sim -\Delta + (C \times \{P_0\}) + (\{P_0\} \times C)$.*

We need some preparation. The first is to relate the divisors Θ and $\Theta^- := [-1]^*\Theta$. Let K_C be an effective canonical divisor. Then $\deg K_C = 2g - 2$ by Corollary 4.1.23, and K_C can be obtained from a point in C^{2g-2}. Let $\kappa := \Phi_{2g-2}(K_C) \in J$ with Φ_{2g-2} defined in (4.5.3). Notice that κ depends only on the divisor class.

Proposition 4.5.12. $\Theta^- = t_\kappa^*\Theta$.

Proof. Let $a \in \Theta$. Then there exists $(P_1, \dots, P_{g-1}) \in C^{g-1}$ such that $a = \Phi_{g-1}(P_1, \dots, P_{g-1})$. Set $D = \sum_{i=1}^{g-1} [P_i] \in \mathrm{Div}(C)$. Then the Riemann–Roch Theorem (Theorem 4.1.22) implies

$$\ell(K_C - D) = \ell(D).$$

[6] For example we can apply Theorem 4.5.9 3 times to get: $t_{-a_1-a_2+a_3}^*\Theta^- \sim t_{-a_1-a_2}^*\Theta^- + t_{a_3}^*\Theta^- - \Theta^-$, $t_{-a_1-a_2}^*\Theta^- \sim t_{-a_1}^*\Theta^- + t_{-a_2}^*\Theta^- - \Theta^-$, and $2\Theta^- = t_{a_3-a_3}^*\Theta^- + \Theta^- \sim t_{a_3}^*\Theta^- + t_{-a_3}^*\Theta^-$.

Since $k \subseteq L(D)$ (as $D \geq 0$), we then have $\ell(K_C - D) \geq 1$. So there exists $f \in k(C)^*$ such that $D' := K_C - D + \operatorname{div}(f) \geq 0$. As $\deg K_C = 2g-2$ and $\deg D = g-1$, we have $\deg D' = g-1$. As an effective divisor of degree $g-1$, $D' = \sum_{i=1}^{g-1}[P_i']$ for some $(P_1', \ldots, P_{g-1}') \in C^{g-1}$. Thus $a = \Phi_{g-1}(P_1, \ldots, P_{g-1}) = \Phi_{2g-2}(K_C + \operatorname{div}(f)) - \Phi_{g-1}(P_1', \ldots, P_{g-1}') = \kappa - \Phi_{g-1}(P_1', \ldots, P_{g-1}') \in \Theta^- + \kappa$.

Hence $\Theta \subseteq \Theta^- + \kappa$. But both Θ and Θ^- are irreducible subvarieties of dimension $g-1$. So $\Theta = \Theta^- + \kappa$. In terms of divisors, we then have $\Theta^- = t_\kappa^* \Theta$. □

Next we state without proving the Seesaw Principle.

Lemma 4.5.13. *Let X and Y be two algebraic varieties, let $L \in \operatorname{Pic}(X \times Y)$. Define for each $x \in X$ the map $i_x \colon Y \to X \times Y$, $y \mapsto (x, y)$. Let $p_1 \colon X \times Y \to X$ be the natural projection.*

(i) If $i_x^* L$ is the trivial line bundle on Y for all $x \in X$, then there exists $L' \in \operatorname{Pic}(X)$ such that $L \simeq p_1^* L'$.

(ii) If furthermore $L|_{X \times \{y_0\}}$ is trivial for some $y_0 \in Y$, then L is trivial on $X \times Y$.

Proof of Theorem 4.5.11. We are studying divisors on $C \times C$. By the seesaw principle (Lemma 4.5.13), it suffices to prove that these two divisors are linearly equivalent when restricted to each slice $\{P\} \times C$ and $C \times \{P\}$ (for all $P \in C$). By symmetry it suffices to use the slices $\{P\} \times C$ for all $P \in C$. Let $i_P \colon C \to C \times C$ be the map $i_P(Q) = (P, Q)$.

We want to apply Proposition 4.5.10 in the following way. It is known by Algebraic Geometry (looking at dimensions) that Φ_g is surjective. Hence Proposition 4.5.10 is equivalent to: $j_a^* \Theta^- \sim D_a + g[P_0]$ for all $a \in J = \mathrm{Cl}^0(C)$, where $D_a \in \operatorname{Div}(C)$ such that $\operatorname{cl}(D_a) = a$.

It is not hard to check: For each $P \neq P_0$, we have

$$i_P^*(-\Delta + (C \times \{P_0\}) + (\{P_0\} \times C)) = -[P] + [P_0].$$

Denote for simplicity $j = j_{P_0}$. Then it remains to prove

$$i_P^* \circ (j \times j)^* \boldsymbol{\delta} \sim -[P] + [P_0]. \tag{4.5.5}$$

To compute $i_P^*(j \times j)^* \boldsymbol{\delta}$, we compute each term separately. Notice that $p_1 \circ (j \times j) \circ i_P$ is constant and $p_2 \circ (j \times j) \circ i_P = j$. Thus

$$i_P^* \circ (j \times j)^* \circ p_1^* \Theta \sim 0 \quad \text{and} \quad i_P^* \circ (j \times j)^* \circ p_2^* \Theta = j^* \Theta.$$

For each $a \in J$, by Proposition 4.5.12 we have $j_a^* \Theta = j_a^* t_{-\kappa}^* \Theta^- = j_{a-\kappa}^* \Theta^-$. Thus $j^* \Theta \sim j_{-\kappa}^* \Theta^-$, which is linearly equivalent to $-K_C + g[P_0]$ by the reinterpretation of Proposition 4.5.10 above.

Similarly, $(m \circ (j \times j) \circ i_P)(Q) = j(P) + j(Q) = j_{j(P)}(Q)$. So

$$i_P^* \circ (j \times j)^* \circ m^* \Theta = j_{j(P)}^* \Theta \sim g[P_0] - ([P] - [P_0]) - K_C.$$

Thus we obtain (4.5.5) by linearity. □

4.6 Rationality

In this section, we give a brief discussion about rationality.

Through the whole section, we let K be a number field and fix $K \subseteq \overline{\mathbb{Q}} \subseteq \mathbb{C}$.

Let X be a projective algebraic variety defined over K. This means that $X \subseteq \mathbb{P}_K^N$ for some N such that $X = V(f_1, \ldots, f_m)$ with $f_1, \ldots, f_m$ homogeneous polynomials with coefficients in K.

We have a natural inclusion $\mathbb{P}^N(K) \subseteq \mathbb{P}^N(\overline{\mathbb{Q}})$. Denote by

$$X(\overline{\mathbb{Q}}) := \{\mathbf{x} \in \mathbb{P}^N(\overline{\mathbb{Q}}) : f_1(\mathbf{x}) = \cdots = f_m(\mathbf{x}) = 0\}.$$

In other words, $X(\overline{\mathbb{Q}})$ is the zero set of $f_1, \ldots, f_m$ viewed as polynomials with coefficients in $\overline{\mathbb{Q}}$. Any point in $X(\overline{\mathbb{Q}})$ is called a $\overline{\mathbb{Q}}$***-point of*** X.

We will use $X_{\overline{\mathbb{Q}}}$ to denote the subvariety of $\mathbb{P}^N_{\overline{\mathbb{Q}}}$ *viewed as a variety over* $\overline{\mathbb{Q}}$.

Since all coefficients of the f_j's are in K, the Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ acts naturally on $X(\overline{\mathbb{Q}})$. We call $\mathbf{x} \in X(\overline{\mathbb{Q}})$ a K***-point*** if $\mathrm{Gal}(\overline{\mathbb{Q}}/K)\mathbf{x} = \mathbf{x}$, *i.e.* the Galois action fixes the point $\mathbf{x}$. The set of K-points of X is denoted by $X(K)$. It can be shown that

$$X(K) := \{\mathbf{x} \in \mathbb{P}^N(K) : f_1(\mathbf{x}) = \cdots = f_m(\mathbf{x}) = 0\},$$

i.e. each K-point is a K-solution to the system defined by the polynomials $f_1, \ldots, f_m$ (which have coefficients in K).

Let us look at the example of a curve embedded into its Jacobian. Let C be a projective irreducible smooth curve defined over K.

Previously, we constructed the Jacobian $\mathrm{Jac}(C_{\overline{\mathbb{Q}}})$ as $\mathrm{Cl}^0(C_{\overline{\mathbb{Q}}})$. We have seen that $\mathrm{Jac}(C_{\overline{\mathbb{Q}}})$ is a projective variety, *i.e.* $\mathrm{Jac}(C_{\overline{\mathbb{Q}}}) \subseteq \mathbb{P}^N_{\overline{\mathbb{Q}}}$ for some N. It turns out that the natural action of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ on $\mathbb{P}^N_{\overline{\mathbb{Q}}}$ preserves $\mathrm{Jac}(C_{\overline{\mathbb{Q}}})$, *i.e.* for any $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/K)$, we have $\sigma(\mathrm{Jac}(C_{\overline{\mathbb{Q}}})) = \mathrm{Jac}(C_{\overline{\mathbb{Q}}})$. Thus $\mathrm{Jac}(C_{\overline{\mathbb{Q}}})$ is defined over K. We use $\mathrm{Jac}(C)$ to denote the Jacobian of C viewed as a variety defined over K.

Let $P_0 \in C(K)$. Then the Abel–Jacobi embedding $j_{P_0} \colon C \to \mathrm{Jac}(C)$ is a morphism defined over K in the same way as above. Thus $j_{P_0}(C)$ is a subvariety of $\mathrm{Jac}(C)$ defined over K. In particular, we have

$$j_{P_0}(C(K)) \subseteq \mathrm{Jac}(C)(K). \tag{4.6.1}$$

We end this section with the following theorem. Let A be an abelian variety defined over K. Then $A(K)$ has a natural structure of abelian groups.

Theorem 4.6.1 (Mordell–Weil Theorem)**.** *As an abelian group $A(K)$ is finitely generated, i.e. $A(K) \simeq \mathbb{Z}^\rho \oplus (\bigoplus \mathbb{Z}/n_i\mathbb{Z})$ for finitely many integers n_i.*

Chapter 5

Height Machine

5.1 Construction and basic properties of the Height Machine

In this section, we define the height function on projective varieties and the height machine.

Let X be an irreducible *projective* variety defined over $\overline{\mathbb{Q}}$. Denote by $\mathbb{R}^{X(\overline{\mathbb{Q}})}$ the set of functions $X(\overline{\mathbb{Q}}) \to \mathbb{R}$, and by $O(1)$ the subset of bounded functions.

The ***Height Machine*** associates to each line bundle $L \in \mathrm{Pic}(X)$ a unique class of functions $\mathbb{R}^{X(\overline{\mathbb{Q}})}/O(1)$, *i.e.* a map

$$\mathbf{h}_X \colon \mathrm{Pic}(X) \to \mathbb{R}^{X(\overline{\mathbb{Q}})}/O(1), \quad L \mapsto \mathbf{h}_{X,L}. \tag{5.1.1}$$

Let $h_{X,L} \colon X(\overline{\mathbb{Q}}) \to \mathbb{R}$ a representative of the class $\mathbf{h}_{X,L}$; it is called a *height function associated with* (X, L).

Construction 5.1.1. *One can construct $h_{X,L}$ as follows. In each case below, $h_{X,L}$ depends on some extra data and hence is not unique. However, it can be shown that any two choices differ by a bounded functions on $X(\overline{\mathbb{Q}})$, and thus the class of $h_{X,L}$ is well-defined.*

(i) If L is very ample, then the global sections of L give rise to a closed immersion $\iota \colon X \to \mathbb{P}^n$ for some n, such that $\iota^*\mathcal{O}(1) \simeq L$. Set $h_{X,L} = h \circ \iota$, with h the Weil height on $\mathbb{P}^n$ from Definition 1.2.1.

(ii) If L is ample, then $L^{\otimes m}$ is very ample for some $m \gg 1$. Set $h_{X,L} = (1/m)h_{X,L^{\otimes m}}$.

(iii) For an arbitrary L, there exist ample line bundles L_1 and L_2 on X such that $L \simeq L_1 \otimes L_2^{\otimes -1}$; see Corollary 4.4.15. Set $h_{X,L} = h_{X,L_1} - h_{X,L_2}$.

Here is how we will arrange to show that the class of $h_{X,L}$ is well-defined in each one of the cases above. For (i), it follows immediately from the following Lemma 5.1.2. For (ii) and (iii), it will be proved in the course of proving Proposition 5.1.3.(ii).

Lemma 5.1.2. *Assume $\phi \colon X \to \mathbb{P}^n$ and $\psi \colon X \to \mathbb{P}^m$ are two morphisms defined over $\overline{\mathbb{Q}}$ such that $\phi^*\mathcal{O}_{\mathbb{P}^n}(1) \simeq \psi^*\mathcal{O}_{\mathbb{P}^m}(1)$. Then as functions on $X(\overline{\mathbb{Q}})$ we have*

$$h_{\mathbb{P}^n} \circ \phi - h_{\mathbb{P}^m} \circ \psi = O(1)$$

where $h_{\mathbb{P}^n}$ (resp. $h_{\mathbb{P}^m}$) is the Weil height on $\mathbb{P}^n$ (resp. on $\mathbb{P}^m$) from Definition 1.2.1.

This $O(1)$ depends on X, ϕ and ψ, but is independent on the point of $X(\overline{\mathbb{Q}})$.

Proof of Lemma 5.1.2. Denote by $L := \phi^*\mathcal{O}_{\mathbb{P}^n}(1) \simeq \psi^*\mathcal{O}_{\mathbb{P}^m}(1)$ the line bundle on X. Choose a basis $\{h_0, \dots, h_N\}$ of $H^0(X, L)$. Then there are linear combinations

$$f_i = \sum_{j=0}^{N} a_{ij}h_j, 0 \le i \le n,$$
$$g_k = \sum_{j=0}^{N} b_{kj}h_j, 0 \le k \le m,$$

with $a_{ij} \in \overline{\mathbb{Q}}$ and $b_{kj} \in \overline{\mathbb{Q}}$, such that

$$\phi = [f_0 : \cdots : f_n] \quad \text{and} \quad \psi = [g_0 : \cdots : g_m].$$

Set $\lambda := [h_0 : \cdots : h_N] \colon X \to \mathbb{P}^N$; then λ is a closed immersion. The matrix $(a_{ij})_{0\le i\le n,\ 0\le j\le N}$ gives rise to a linear map $A \colon \mathbb{P}^N \to \mathbb{P}^n$, and the matrix $(b_{kj})_{0\le k\le m,\ 0\le j\le N}$ gives rise to a linear map $B \colon \mathbb{P}^N \to \mathbb{P}^m$. Notice $A \circ \lambda = \phi$ and $B \circ \lambda = \psi$. So both A and B are well-defined over $\lambda(X)$. Hence we can apply Theorem 1.2.15 and obtain

$$h(\phi(x)) = h(A(\lambda(x))) = h(\lambda(x)) + O(1) \text{ and } h(\psi(x)) = h(B(\lambda(x))) = h(\lambda(x)) + O(1)$$

for all $x \in X(\overline{\mathbb{Q}})$. Taking the difference of these two equalities, we get the desired equality. □

Here are some basic properties of the Height Machine. These properties, or more precisely properties (i)–(iii), also uniquely determine (5.1.1).

Proposition 5.1.3. *We have*

(i) (Normalization) Let h be the Weil height from Definition 1.2.1. Then for all $\mathbf{x} \in \mathbb{P}^n(\overline{\mathbb{Q}})$, we have

$$h_{\mathbb{P}^n,\mathcal{O}(1)}(\mathbf{x}) = h(\mathbf{x}) + O(1).$$

(ii) (Additivity) Let L and M be two line bundles on X. Then for all $x \in X(\overline{\mathbb{Q}})$, we have

$$h_{X,L\otimes M}(x) = h_{X,L}(x) + h_{X,M}(x) + O(1).$$

(iii) (Functoriality) Let $\phi \colon X \to Y$ be a morphism of irreducible projective varieties and let L be a line bundle on Y. Then for all $x \in X(\overline{\mathbb{Q}})$, we have

$$h_{X,\phi^*L}(x) = h_{Y,L}(\phi(x)) + O(1).$$

(iv) (Positivity) If $s \in H^0(X, L)$ is a global section, then for all $x \in (X \setminus \operatorname{div}(s))(\overline{\mathbb{Q}})$ we have

$$h_{X,L}(x) \ge O(1).$$

(v) (Northcott property) Assume L is ample. Let K_0 be a number field on which X is defined. Then for any $d \ge 1$ and any constant B, the set

$$\{x \in X(K) : [K : K_0] \le d,\ h_{X,L}(x) \le B\}$$

is a finite set.

The $O(1)$'s that appear in the proposition depend on the varieties, line bundles, morphisms, and the choices of the representatives in the classes of height functions. But they are independent of the points on the varieties.

Proof of Proposition 5.1.3. Part (i) follows from the definition and the fact that $x_0, \ldots, x_n$ is a basis of $H^0(\mathbb{P}^n, \mathcal{O}(1))$. Notice that Lemma 5.1.2 is implicitly used.

Next we check (ii). We start with the case where both L and M are very ample. Then the global sections of L (resp. of M) give rise to a closed immersion $\phi_L \colon X \to \mathbb{P}^n$ (resp. $\psi \colon X \to \mathbb{P}^m$). Composing with the Segre embedding $S_{n,m} \colon \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^N$ (with $N = (n+1)(m+1)-1$) from (1.2.5), we obtain

$$\phi_L \otimes \phi_M \colon X \to \mathbb{P}^N, \quad x \mapsto \phi_L(x) \otimes \phi_M(x).$$

Recall that $S_{n,m}^* \mathcal{O}_{\mathbb{P}^N}(1) \simeq \mathcal{O}(1,1)$ by Lemma 4.4.10. So $(\phi_L \otimes \phi_M)^* \mathcal{O}_{\mathbb{P}^N}(1) \simeq L \otimes M$. So $h_{X,L\otimes M}(x) = h_{\mathbb{P}^N}(\phi_L(x) \otimes \phi_M(x))$, which equals $h_{\mathbb{P}^n}(\phi_L(x)) + h_{\mathbb{P}^m}(\phi_M(x))$ by Proposition 1.2.14.(i), and hence equals $h_{X,L}(x) + h_{X,M}(x) + O(1)$.

At this stage, we are ready to establish case (ii) of Construction 5.1.1. Suppose L is ample. If m and n satisfy that $L^{\otimes m}$ and $L^{\otimes n}$ are very ample, then $L^{\otimes mn}$ is very ample. Apply Proposition 5.1.3.(ii) to $L^{\otimes m}$ (n times), then we get $h_{X,L^{\otimes mn}} = n h_{X,L^{\otimes m}} + O(1)$. Similarly (apply Proposition 5.1.3.(ii) to $L^{\otimes n}$ (m times)) we have $h_{X,L^{\otimes mn}} = m h_{X,L^{\otimes n}} + O(1)$. Thus up to $O(1)$, we have $\frac{1}{m} h_{X,L^{\otimes m}} = \frac{1}{n} h_{X,L^{\otimes n}}$. Hence $h_{X,L}$ is well-defined up to $O(1)$ if L is ample.

Now Proposition 5.1.3.(ii) for the case where both L and M are ample follows from the very ample case and the definition of the height function in this case.

For arbitrary L and M, write $L = L_1 \otimes L_2^{\otimes -1}$ and $M = M_1 \otimes M_2^{\otimes -1}$ with L_1, L_2, M_1 and M_2 ample. Then $L_1 \otimes M_1$ and $L_2 \otimes M_2$ are ample line bundles on X, with $L \otimes M \simeq (L_1 \otimes M_1) \otimes (L_2 \otimes M_2)^{\otimes -1}$. Thus up to $O(1)$, we have

$$h_{X,L\otimes M} = h_{X,L_1\otimes M_1} - h_{X,L_2\otimes M_2} = h_{X,L_1} + h_{X,M_1} - h_{X,L_2} - h_{X,M_2} = h_{X,L} + h_{X,M}.$$

Notice that this also establishes case (iii) of Construction 5.1.1 (that $h_{X,L}$ is well-defined up to $O(1)$ for an arbitrary L).

For (iii): By (ii) it suffices to prove the assertion for L very ample. Let $\iota_L \colon Y \to \mathbb{P}^n$ be a closed immersion given by global sections of L; then $\iota_L^* \mathcal{O}(1) \simeq L$. In particular, $h_{\mathbb{P}^n} \circ \iota_L = h_{Y,L} + O_Y(1)$ by part (i). There exists some very ample M on X such that $\phi^* L \otimes M$ is very ample; see Proposition 4.4.14. The global sections of M give rise to a closed immersion $\iota_M \colon X \to \mathbb{P}^m$. Hence we have a morphism $(\iota_L \circ \phi, \iota_M) \colon X \to \mathbb{P}^n \times \mathbb{P}^m$, which composed with the Segre embedding gives a closed immersion $\iota \colon X \to \mathbb{P}^N$. One can check that $\iota^* \mathcal{O}(1) \simeq \phi^* L \otimes M$. So as in the proof of part (ii), we have up to $O_X(1)$

$$h_{X,\phi^*L\otimes M} = h_{\mathbb{P}^N} \circ \iota = h_{\mathbb{P}^n} \circ \iota_L \circ \phi + h_{\mathbb{P}^m} \circ \iota_M = h_{Y,L} \circ \phi + h_{X,M}.$$

Hence we are done by part (ii).

For (iv): There exist a positive integer k and a very ample line bundle M on X such that $L^{\otimes k} \otimes M$ is very ample on X; see Proposition 4.4.14. Notice that $s^k \in H^0(X, L^{\otimes k})$. Let $\{f_0, \ldots, f_m\}$ be a basis of $H^0(X, M)$; then we have a closed immersion $\iota_M := [f_0 : \cdots : f_m] \colon X \to \mathbb{P}^m$. One can complete $s^k f_0, \ldots, s^k f_m$ to a basis $\{s^k f_j, g_i\}_{0\le j\le m, 1\le i\le n}$ of $H^0(X, L^{\otimes k} \otimes M)$, and thus obtain a closed immersion $\iota \colon X \to \mathbb{P}^N$. Now up to $O(1)$, $h_{X,L^{\otimes k}} = h_{\mathbb{P}^N} \circ \iota - h_{\mathbb{P}^m} \circ \iota_M$ by part (ii). For any $x \in (X \setminus \operatorname{div}(s))(\overline{\mathbb{Q}})$, we have $\iota_M(x) = [f_0(x) : \cdots : f_m(x)] = [s(x)^k f_0(x) : \cdots : s(x)^k f_m(x)] \in \mathbb{P}^m(\overline{\mathbb{Q}})$, and so

$$h_{\mathbb{P}^N} \circ \iota(x) - h_{\mathbb{P}^m} \circ \iota_M(x) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in M_K} \left(\log \max \left\{ \max_j \|s(x)^k f_j(x)\|_v, \max_i \|g_i(x)\|_v \right\} - \log \max_j \|s(x)^k f_j(x)\|_v \right)$$

for an appropriate number field K, and hence is ≥ 0. Hence we are done.

For (v), it suffices to prove for L very ample. Then the conclusion follows immediately from the Northcott Property for Weil height (Theorem 1.2.5). □

5.2 Normalized Height after Néron and Tate

Let X be an irreducible projective variety defined over $\overline{\mathbb{Q}}$.

The Height Machine associates to each line bundle $L \in \mathrm{Pic}(X)$ a height function $h_L \colon X(\overline{\mathbb{Q}}) \to \mathbb{R}$. However, these height functions are well-defined only up to $O(1)$. It is sometimes desirable to find particular representatives.

While one can always fix a representative by fixing every operation needed to define h_L (for example, the basis of $H^0(X, L)$ giving the embedding of X into some $\mathbb{P}^N$ if L is very ample), for some particular (X, L) we have some more canonical choices. In this section, we discuss one case developed by Néron and Tate.

Assume that $\phi \colon X \to X$ is a morphism satisfying $\phi^* L \simeq L^{\otimes \alpha}$ for some integer $\alpha > 1$.

Theorem 5.2.1. *There exists a unique height function*

$$\hat{h}_{X,\phi,L} \colon X(\overline{\mathbb{Q}}) \to \mathbb{R}$$

with the following properties.

(i) $\hat{h}_{X,\phi,L}(x) = h_{X,L}(x) + O(1)$ *for all* $x \in X(\overline{\mathbb{Q}})$,

(ii) $\hat{h}_{X,\phi,L}(\phi(x)) = \alpha \hat{h}_{X,\phi,L}(x)$ *for all* $x \in X(\overline{\mathbb{Q}})$.

The height function $\hat{h}_{X,\phi,L}$ *depends only on the isomorphism class of* L. *Moreover, it can be computed as the limit*

$$\hat{h}_{X,\phi,L}(x) = \lim_{n \to \infty} \frac{1}{\alpha^n} h_{X,L}(\phi^n(x)) \tag{5.2.1}$$

with ϕ^n *the* n*-fold iterate of* ϕ.

Property (i) says that $\hat{h}_{X,\phi,L}$ is in the class of heights of $h_{X,L}$. The height function is sometimes called the *canonical height function*.

Here is an example of the application of Theorem 5.2.1. Let $\phi \colon \mathbb{P}^n \to \mathbb{P}^n$ be given by homogeneous polynomials of degree $d > 1$, then $\phi^* \mathcal{O}(1) \simeq \mathcal{O}(d) = \mathcal{O}(1)^{\otimes d}$. If $\phi([x_0 : \cdots : x_n]) = [x_0^d : \cdots : x_n^d]$, then one can check that $\hat{h}_{\mathbb{P}^n,\phi,\mathcal{O}(1)}$ is precisely the Weil height.

A more important example for the Tate Limit Process (5.2.1) is the definition of the *Néron–Tate heights on abelian varieties*. This height turns out to be extremely useful. We will come back to this in the next section.

Before moving on to the proof, let us have a digest. The morphism ϕ induces a $\mathbb{Z}$-linear map $\phi^* \colon \mathrm{Pic}(X) \to \mathrm{Pic}(X)$.[1] Tensoring with $\mathbb{R}$ gives a linear map $\phi^* \colon \mathrm{Pic}(X) \otimes_{\mathbb{Z}} \mathbb{R} \to \mathrm{Pic}(X) \otimes_{\mathbb{Z}} \mathbb{R}$ of real vector spaces of finite dimension. Say L is non-trivial. Then the assumption $\phi^* L \simeq L^{\otimes \alpha}$ implies that L is an eigenvector for the eigenvalue α. The assumption $\alpha > 1$ guarantees that the *Tate Limit Process* (5.2.1) will work in the end.

Proof of Theorem 5.2.1. Applying Proposition 5.1.3.(iii) to the relation $\phi^* L \simeq L^{\otimes \alpha}$, we get a constant C such that

$$|h_{X,L}(\phi(y)) - \alpha h_{X,L}(y)| \leq C \quad \text{for all } y \in X(\overline{\mathbb{Q}}).$$

[1] The "addition" on the group $\mathrm{Pic}(X)$ is $\otimes$.

Notice that C depends on X, L, ϕ and the choice of the height function $h_{X,L}$.

Claim: For any $x \in X(\overline{\mathbb{Q}})$, the sequence $\alpha^{-n} h_{X,L}(\phi^n(x))$ converges.

We prove this by Cauchy. The proof uses the telescoping sum. Let $n \geq m$ and compute

$$\begin{aligned}
\left|\alpha^{-n} h_{X,L}(\phi^n(x)) - \alpha^{-m} h_{X,L}(\phi^m(x))\right| &= \left|\sum_{i=m+1}^{n} \alpha^{-i} \left(h_{X,L}(\phi^i(x)) - \alpha h_{X,L}(\phi^{i-1}(x))\right)\right| \text{ (telescoping sum)} \\
&\leq \sum_{i=m+1}^{n} \alpha^{-i} \left|h_{X,L}(\phi^i(x)) - \alpha h_{X,L}(\phi^{i-1}(x))\right| \text{ (triangle inequality)} \\
&\leq \sum_{i=m+1}^{n} \alpha^{-i} C \quad \text{from above with } y = \phi^{i-1}(x).
\end{aligned}$$

So

$$\left|\alpha^{-n} h_{X,L}(\phi^n(x)) - \alpha^{-m} h_{X,L}(\phi^m(x))\right| \leq \frac{\alpha^{-m} - \alpha^{-n}}{\alpha - 1} C. \tag{5.2.2}$$

But $\frac{\alpha^{-m}-\alpha^{-n}}{\alpha-1} C \to 0$ as $n > m \to \infty$. Thus the sequence $\alpha^{-n} h_{X,L}(\phi^n(x))$ is Cauchy, and hence converges. So we can define $\hat{h}_{X,\phi,L}(x)$ as in (5.2.1).

Now we verify the properties (i) and (ii). For (i), take $m = 0$ and let $n \to \infty$ in the inequality (5.2.2). We then get

$$\left|\hat{h}_{X,\phi,L}(x) - h_{X,L}(x)\right| \leq \frac{C}{\alpha - 1}. \tag{5.2.3}$$

And this gives (a more explicit form of) property (i).

Property (ii) follows directly from the computation

$$\begin{aligned}
\hat{h}_{X,\phi,L}(\phi(x)) &= \lim_{n\to\infty} \frac{1}{\alpha^n} h_{X,L}\left(\phi^n(\phi(x))\right) \\
&= \lim_{n\to\infty} \frac{\alpha}{\alpha^{n+1}} h_{X,L}(\phi^{n+1}(x)) \\
&= \alpha \hat{h}_{X,\phi,L}(x).
\end{aligned}$$

It remains to prove the uniqueness. Suppose $\hat{h}$ and $\hat{h}'$ are two functions with properties (i) and (ii). Set $g := \hat{h} - \hat{h}'$. Then (i) implies that g is bounded, say $|g(x)| \leq C'$ for all $x \in X(\overline{\mathbb{Q}})$. Property (ii) implies that $g \circ \phi = \alpha g$ and thus $g \circ \phi^n = \alpha^n g$ for all $n \geq 1$. Hence

$$|g(x)| = \frac{|g(\phi^n(x))|}{\alpha^n} \leq \frac{C'}{\alpha^n} \xrightarrow{n\to\infty} 0.$$

Thus $g \equiv 0$ and hence $\hat{h} = \hat{h}'$. We are done. □

Proposition 5.2.2. *Assume furthermore that L is ample. Then*

(i) $\hat{h}_{X,\phi,L}(x) \geq 0$ for all $x \in X(\overline{\mathbb{Q}})$;

*(ii) $\hat{h}_{X,\phi,L}(x) = 0$ if and only if x is **preperiodic** for ϕ, i.e. $O_\phi^+(x) := \{x, \phi(x), \phi^2(x), \ldots\}$ is a finite set.*

Proof. For (i): As L is ample, $L^{\otimes m}$ is very ample for some $m \gg 1$. Take a basis $\{s_1, \ldots, s_k\}$ of $H^0(X, L^{\otimes m})$, then $\bigcap_{i=1}^k \operatorname{div}(s_i) = \emptyset$. By Proposition 5.1.3.(iv) applied to each s_i, we can choose a representative $h_{X,L^{\otimes m}}$ with $h_{X,L^{\otimes m}}(x) \geq 0$ for all $x \in X(\overline{\mathbb{Q}})$. Thus $h_{X,L}(x) = (1/m) h_{X,L^{\otimes m}}(x) \geq 0$ for all $x \in X(\overline{\mathbb{Q}})$. So $\hat{h}_{X,\phi,L}(x) \geq 0$ for all $x \in X(\overline{\mathbb{Q}})$ by (5.2.1).

Let us prove property (ii). Take $x \in X(\overline{\mathbb{Q}})$. For $\Leftarrow$: It is clear that $h_{X,L}(\phi^n(x))$ is bounded because $O_\phi^+(x)$ is a finite set. So $\alpha^{-n} h_{X,L}(\phi^n(x)) \to 0$ as $n \to \infty$. Thus $\hat{h}_{X,\phi,L}(x) = 0$ by (5.2.1).

It remains to prove $\Rightarrow$ of property (ii). Take a number field K such that X, L, ϕ are defined over K and $x \in X(K)$. Suppose $\hat{h}_{X,\phi,L}(x) = 0$. Then for any $n \geq 1$, we have

$$h_{X,L}(\phi^n(x)) = \hat{h}_{X,\phi,L}(\phi^n(x)) + O(1) = \alpha^n \hat{h}_{X,\phi,L}(x) + O(1) = O(1).$$

Here the constant $O(1)$ depends only on X and L. As all $\phi^n(x)$ are in $X(K)$, we obtain a constant B such that

$$O_\phi^+(x) \subseteq \{y \in X(K) : h_{X,L}(y) \leq B\}.$$

Thus $O_\phi^+(x)$ is a finite set by the Northcott property (Proposition 5.1.3.(v)). We are done. □

This proposition is important when we study the canonical heights on abelian varieties in the next section.

Here is an application.

Corollary 5.2.3 (Kronecker's Theorem)**.** *Consider the Weil height h on $\overline{\mathbb{Q}} = \mathbb{A}^1(\overline{\mathbb{Q}})$. Let $\zeta \in \overline{\mathbb{Q}}^*$. Then $h(\zeta) = 0$ if and only if ζ is a root of unity.*

Proof. Consider the morphism $\phi\colon \mathbb{P}^1 \to \mathbb{P}^1$, $[x_0 : x_1] \mapsto [x_0^2 : x_1^2]$. Then $h(x) = \hat{h}_{\mathbb{P}^1,\phi,\mathcal{O}(1)}([1 : x])$ for all $x \in \overline{\mathbb{Q}}$. For $\Rightarrow$, suppose $h(\zeta) = 0$. By Proposition 5.2.2.(ii), $\{[1 : \zeta], [1 : \zeta^2], [1 : \zeta^4], \ldots\}$ is a finite set. So $\zeta^{2^i} = \zeta^{2^j}$ for some $i \neq j$. Thus ζ is a root of unity. For $\Leftarrow$, suppose $\zeta^n = 1$. Fermat's Little Theorem implies $2^{\phi(n)} \equiv 1 \pmod{n}$ for the Euler-ϕ function. Thus $\{[1 : \zeta], [1 : \zeta^2], [1 : \zeta^4], \ldots\}$ is a finite set, and hence $h(\zeta) = \hat{h}_{\mathbb{P}^1,\phi,\mathcal{O}(1)}([1 : \zeta]) = 0$ by Proposition 5.2.2.(ii). □

5.3 Néron–Tate height on abelian varieties

In this section, we discuss about normalized height functions on abelian varieties.

Let A be an abelian variety defined over $\overline{\mathbb{Q}}$. Let $L \in \mathrm{Pic}(A)$ be a line bundle such that $L \simeq [-1]^* L$ (we call such an L ***even***). By Corollary 4.5.8, we have

$$[n]^* L \simeq L^{\otimes n^2} \tag{5.3.1}$$

for all $n \in \mathbb{Z}$.

Let us apply Theorem 5.2.1 to $[2]\colon A \to A$ and L. Then we obtain the normalized height function

$$\hat{h}_{A,L}\colon A(\overline{\mathbb{Q}}) \to \mathbb{R}. \tag{5.3.2}$$

This function is called the ***Néron–Tate height*** on A with respect to L. Compared to the notation in the last section, we omitted the map [2] in the subscript. This is justified by the following proposition, which implies that we can replace [2] by any $[n]$ with $n \geq 2$ in the definition of $\hat{h}_{A,L}$.

Proposition 5.3.1. *For each $N \in \mathbb{Z}$, we have $\hat{h}_{A,L}([N]x) = N^2 \hat{h}_{A,L}(x)$ for all $x \in A(\overline{\mathbb{Q}})$. In particular, we have*

$$\hat{h}_{A,L}(x) = \lim_{N \to \infty} \frac{h_{A,L}([N]x)}{N^2}.$$

Proof. We have $[N]^* L \simeq L^{\otimes N^2}$ by (5.3.1). Thus (ii) and (iii) of Proposition 5.1.3 (applied to the height function $\hat{h}$) yield $\hat{h}_{A,L}([N]y) = \hat{h}_{A,[N]^*L}(y) + O(1) = \hat{h}_{A,L^{\otimes N^2}}(y) + O(1) = N^2 \hat{h}_{A,L}(y) + O(1)$ for all $y \in A(\overline{\mathbb{Q}})$, where $O(1)$ is a constant depending on A and L. In particular let $y = [2^n]x$, then we have

$$\hat{h}_{A,L}([2^n][N]x) = N^2 \hat{h}_{A,L}([2^n]x) + O(1) = N^2 4^n \hat{h}_{A,L}(x) + O(1)$$

where the last equality follows from Theorem 5.2.1.(ii). Dividing both sides by 4^n and letting $n \to \infty$, we get $\hat{h}_{A,L}([N]x) = N^2 \hat{h}_{A,L}(x)$.

For the "In particular" part, we know (Theorem 5.2.1.(i)) that $\hat{h}_{A,L} = h_{A,L} + O(1)$. Thus

$$\lim_{N\to\infty} \frac{h_{A,L}([N]x)}{N^2} = \lim_{N\to\infty} \frac{\hat{h}_{A,L}([N]x) + O(1)}{N^2} = \hat{h}_{A,L}(x).$$

We are done. □

Proposition 5.3.2. *Assume L is ample. Then*

(i) $\hat{h}_{A,L}(x) \geq 0$ for all $x \in A(\overline{\mathbb{Q}})$;

(ii) $\hat{h}_{A,L}(x) = 0$ if and only if x is a torsion point, i.e. $[N]x = 0$ for some integer $N \neq 0$;

Proof. Part (i) follows immediately from Proposition 5.2.2.(i).

For (ii), we use Proposition 5.2.2.(ii). Assume $\hat{h}_{A,L}(x) = 0$. Then $\{[2^n]x : n \geq 1\}$ is a finite set by Proposition 5.2.2.(ii). Thus $[2^n]x = [2^m]x$ for some $m > n$. Thus $[2^m - 2^n]x = 0$ and $2^m - 2^n \neq 0$, and hence x is a torsion point. Conversely assume $[N]x = 0$ with $N \neq 0$. Then the set $O^+_{[N]}(x) := \{x, [N]x, [N^2]x, \cdots\}$ is a finite set. So Proposition 5.2.2.(ii) implies that $\hat{h}_{A,[N],L}(x) = 0$. But $\hat{h}_{A,[N],L} = \hat{h}_{A,L}$ by Proposition 5.3.1. Hence we are done. □

We finish this section by the following discussion.

Take a finitely generated subgroup Γ of $A(\overline{\mathbb{Q}})$. By linearity, the Néron–Tate height $\hat{h}_{A,L}$ extends to a function $\Gamma_{\mathbb{R}} := \Gamma \otimes_{\mathbb{Z}} \mathbb{R} \to \mathbb{R}$. By abuse of notation we still denote this function by $\hat{h}_{A,L}$.

Proposition 5.3.3. *For each finitely generated subgroup Γ of $A(\overline{\mathbb{Q}})$, $\hat{h}_{A,L}$ is a quadratic form on $\Gamma_{\mathbb{R}}$ which is furthermore positive definite.*

Proof. In view of Proposition 5.3.2.(i), in order to prove that $\hat{h}_{A,L}$ is a quadratic form on $A(\overline{\mathbb{Q}})$, it suffices to show that the pairing

$$\langle \cdot, \cdot \rangle_L \colon A(\overline{\mathbb{Q}}) \times A(\overline{\mathbb{Q}}) \to \mathbb{R}, \quad (a, b) \mapsto \frac{1}{2}\left(\hat{h}_{A,L}(a+b) - \hat{h}_{A,L}(a) - \hat{h}_{A,L}(b)\right) \tag{5.3.3}$$

is bilinear. This easily follows from the theorem of the square (Theorem 4.5.9) because $\hat{h}_{A,L}(x) = \hat{h}_{A,t_x^* L}(0)$ for all $x \in A(\overline{\mathbb{Q}})$.

Notice that $\hat{h}_{A,L}$ is then a quadratic form on $\Gamma_{\mathbb{R}}$ by linearity.

To show that $\hat{h}_{A,L}$ is positive definite on $\Gamma_{\mathbb{R}}$, we need to prove two things by Lemma 5.3.4. In order to distinguish $\hat{h}_{A,L}$ on Γ and on $\Gamma_{\mathbb{R}}$, we denote the latter by q. We use $\overline{\Gamma}$ to denote the image of $\Gamma \to \Gamma_{\mathbb{R}}$; it is isomorphic to Γ mod the torsion points.

(a) If $0 \neq \gamma \in \Gamma_{\mathbb{R}}$ lies in $\overline{\Gamma}$, then $q(\gamma) > 0$.

(b) For every $C > 0$, the set $\{\gamma \in \overline{\Gamma} : q(\gamma) \leq C\}$ is finite.

For (a), it easily follows from (i) and (ii) of the current proposition. For (b), suppose γ is the image of some $x \in \Gamma$. Then $q(\gamma) \leq C \Rightarrow \hat{h}_{A,L}(x) \leq C$. As Γ is finitely generated, there exists a number field K such that $\Gamma \subseteq A(K)$. Thus we are looking at $\{x \in A(K) : \hat{h}_{A,L}(x) \leq C\}$, which is a finite set by the Northcott property (Proposition 5.1.3.(v)). So (b) is also established. We are done. □

Lemma 5.3.4. *Let M be a finitely generated abelian group and let $q\colon M \to \mathbb{R}$ be a quadratic form. Set $q_{\mathbb{R}}\colon M_{\mathbb{R}} := M \otimes_{\mathbb{Z}} \mathbb{R} \to \mathbb{R}$ to be the quadratic form defined by linearity. Then $q_{\mathbb{R}}$ is positive definite if and only if the following two conditions are satisfied:*

(a) $q(x) > 0$ for all $x \in \overline{M} \setminus \{0\}$, where $\overline{M}$ is the image of $M \to M_{\mathbb{R}}$;

(b) For every $C > 0$, the set $\{x \in \overline{M} : q_{\mathbb{R}}(x) \leq C\}$ is finite.

Part (b) is necessary as is shown by the following example. Suppose α is a transcendental number in $\mathbb{R}$, then the quadratic form in $\mathbb{R}^2$ given by $q(x_1, x_2) := (x_1 - \alpha x_2)^2$ is not positive definite since $q(\alpha, 1) = 0$, but $q(x_1, x_2) > 0$ for all $(x_1, x_2) \in \overline{\mathbb{Q}}^2 \setminus \{0\}$!

Proof. The direction $\Rightarrow$ is easy. We prove $\Leftarrow$. Assume $q_{\mathbb{R}}$ is not positive definite. Then there exists $y \in M_{\mathbb{R}} \setminus \{0\}$ such that $q_{\mathbb{R}}(y) = 0$.

We claim that $y \notin \overline{M}_{\mathbb{Q}} = M_{\mathbb{Q}}$. Indeed if $y \in \overline{M}_{\mathbb{Q}}$, then $Ny \in \overline{M} \setminus \{0\}$ for some $0 \neq N \in \mathbb{N}$. Then $q(Ny) > 0$ by (a). But q is quadratic, so $q(Ny) = N^2 q(y) > 0$. This contradicts the choice of y.

Choose a basis $\{x_1, \ldots, x_r\}$ of $\overline{M}$; it is also a basis of $M_{\mathbb{R}}$. For any $n \in \mathbb{N}$, there exists $y_n \in \overline{M}$ such that the coordinates of $y_n - ny$ are in the interval $[0, 1]$. Thus $y_n - ny$ is contained in the compact cube $\{\sum_{i=1}^{r} \alpha_i x_i : 0 \leq \alpha_i \leq 1\}$. But $q_{\mathbb{R}}(y_n) = q_{\mathbb{R}}(y_n - ny)$ (since $q_{\mathbb{R}}(y) = 0$)[2] and hence is bounded on the cube, say by C. Since $y \notin \overline{M}_{\mathbb{Q}}$, the set $\{y_n : n \in \mathbb{N}\}$ is infinite and is contained in $\{x \in \overline{M} : q_{\mathbb{R}}(x) \leq C\}$. This contradicts (b). Hence we are done. □

[2]This can be seen from (for example) the bilinear pairing associated with the quadratic form $q_{\mathbb{R}}$.

Chapter 6

Mordell Conjecture

6.1 Statement and Gap Principle

The goal of this chapter is to prove the famous *Falting's Theorem*, also known as the *Mordell Conjecture*.

Theorem 6.1.1. *Let C be an irreducible projective smooth curve defined over a number field K. If the genus of C (denoted by g) is ≥ 2, then $C(K)$ is a finite set.*

This is a very strong result. The genus g is a topological invariant of C, and the set of rational points $C(K)$ is arithmetic information.

For the proof, we will follow Vojta's approach and take Bombieri's simplification.

Let $J = \mathrm{Jac}(C)$ be the Jacobian of C. Fix $P_0 \in C(K)$, and let

$$j = j_{P_0} \colon C \to J, \quad P \mapsto \mathrm{cl}([P] - [P_0])$$

be the Abel–Jacobi embedding via P_0. Then j is defined over K. Notice that at this step, we assumed $C(K) \neq \emptyset$. This does not cause problem for the purpose of proving Mordell's Conjecture, because otherwise we simply have $\#C(K) = 0$.

6.1.1 Mumford's Formula

Let $\Theta := j_{P_0}(C) + \cdots + j_{P_0}(C)$ ($g-1$ copies). Then it has dimension $g-1$, and hence is a Weil divisor on J. Denote by $\Theta^- := [-1]^*\Theta$; as a variety it is $-j_{P_0}(C) - \cdots - j_{P_0}(C)$ ($g-1$ copies). Both Θ and Θ^- are ample divisors; see Theorem 4.4.16.

Lemma 6.1.2. *We have:*

(i) $[2]^*\Theta = 3\Theta + \Theta^-$;

(ii) As divisors on C, we have $j^*\Theta^- \sim g[P_0]$.

Proof. Part (i) follows immediately from Corollary 4.5.8 applied to $n = 2$ and $L = \mathcal{O}(\Theta)$. Part (ii) follows immediately from Proposition 4.5.10 applied to the point $(P_0, \ldots, P_0) \in C^g$, because $j(P_0) = 0$. □

In practice, it is more natural to work with symmetric line bundles or symmetric divisors on J. Thus we consider the divisor $\Theta + \Theta^-$, which clearly satisfies $[-1]^*(\Theta + \Theta^-) = \Theta + \Theta^-$. It is ample. Thus for the associated line bundle $\mathcal{O}(\Theta + \Theta^-)$, the associated Néron–Tate height

$\hat{h}_{\mathcal{O}(\Theta+\Theta^-)}\colon J(\overline{\mathbb{Q}}) \to \mathbb{R}_{\geq 0}$ is a quadratic form, which by Proposition 5.3.3 extends linearly to a *positive definite quadratic form*[1]

$$\hat{h}_{\Theta+\Theta^-}\colon J(K)_{\mathbb{R}} := J(K) \otimes_{\mathbb{Z}} \mathbb{R} \to \mathbb{R}_{\geq 0}.$$

Then $J(K)_{\mathbb{R}}$, which is a finite-dimensional $\mathbb{R}$-vector space by the Mordell–Weil Theorem (Theorem 4.6.1), is endowed with a norm given by $|\cdot| := \hat{h}^{1/2}_{\Theta+\Theta^-}$. We can then define the *angle* between each $z, w \in J(K)_{\mathbb{R}}$. More precisely, let $\langle\cdot,\cdot\rangle\colon J(K)_{\mathbb{R}} \times J(K)_{\mathbb{R}} \to \mathbb{R}$ be the bilinear pairing associated with $\hat{h}_{\Theta+\Theta^-}$, *i.e.*

$$\langle z, w\rangle = \frac{1}{2}\left(\hat{h}_{\Theta+\Theta^-}(z+w) - \hat{h}_{\Theta+\Theta^-}(z) - \hat{h}_{\Theta+\Theta^-}(w)\right). \tag{6.1.1}$$

Then the angle α between z and w is given by $\cos\alpha = \langle z, w\rangle/|z||w|$.

The bilinear pairing (6.1.1) can be seen as a height on $J \times J$ as follows. Recall the Poincaré divisor $\boldsymbol{\delta} := m^*\Theta - p_1^*\Theta - p_2^*\Theta$ on $J \times J$, where p_i is the natural projection to the i-th factor and m is the addition

$$\begin{array}{ccc} J \times J & \xrightarrow{p_1} & J \\ {\scriptstyle p_2}\downarrow & \searrow^{m} & \\ J & & J. \end{array}$$

Lemma 6.1.3. *We have $\hat{h}_{\boldsymbol{\delta}}(a,a) = \hat{h}_{\Theta+\Theta^-}(a)$ for all $a \in J(\overline{\mathbb{Q}})\otimes\mathbb{R}$. As a consequence, $\hat{h}_{\boldsymbol{\delta}}(a,b) = \langle a, b\rangle$ for all $a, b \in J(K)_{\mathbb{R}}$.*

Proof. For the diagonal embedding $\Delta_J\colon J \to J \times J$, we have

$$\Delta_J^*\boldsymbol{\delta} = (m\circ\Delta_J)^*\Theta - (p_1\circ\Delta_J)^*\Theta - (p_2\circ\Delta_J)^*\Theta = [2]^*\Theta - 2\Theta = \Theta + \Theta^-$$

where the last equality follows from Lemma 6.1.2.(i). So

$$\hat{h}_{\boldsymbol{\delta}}(a,a) = \hat{h}_{\boldsymbol{\delta}}(\Delta_J(a)) = \hat{h}_{\Delta_J^*\boldsymbol{\delta}}(a) = \hat{h}_{\Theta+\Theta^-}(a).$$

And $\hat{h}_{\boldsymbol{\delta}}(a,b) = \langle a, b\rangle$ since $\hat{h}_{\boldsymbol{\delta}}(a+b, a+b) = \hat{h}_{\boldsymbol{\delta}}(a,a) + \hat{h}_{\boldsymbol{\delta}}(b,b) + 2\hat{h}_{\boldsymbol{\delta}}(a,b)$. □

Now we are ready to prove Mumford's Formula. Before doing this, recall that

$$(j\times j)^*\boldsymbol{\delta} \sim -\Delta + (C\times\{P_0\}) + (\{P_0\}\times C) \tag{6.1.2}$$

as divisors on $C \times C$ by Theorem 4.5.11. Here Δ is the diagonal on $C \times C$.

Proposition 6.1.4 (Mumford's Formula)**.** *We have, for all $P, Q \in C(\overline{\mathbb{Q}})$,*

$$h_{C\times C,\Delta}(P,Q) = \frac{1}{2g}|j(P)|^2 + \frac{1}{2g}|j(Q)|^2 - \langle j(P), j(Q)\rangle + O(|j(P)| + |j(Q)| + 1). \tag{6.1.3}$$

[1] By Theorem 4.4.7, we are allowed to identify (Cartier) divisor classes and isomorphism classes of line bundles. So for each (Cartier) divisor D, we use h_D to denote the height function $h_{\mathcal{O}(D)}$.

Proof. Denote by $z = j(P)$ and $w = j(Q)$. We have

$$\begin{aligned}
\langle z, w\rangle &= \hat{h}_{J\times J,\boldsymbol{\delta}}(z,w) \qquad \text{by Lemma 6.1.3}\\
&= h_{C\times C,(j\times j)^*\boldsymbol{\delta}}(P,Q) + O(1) \qquad \text{by Proposition 5.1.3.(iii)}\\
&= h_{C\times C,\mathcal{O}(\{P_0\}\times C + C\times\{P_0\}-\Delta)}(P,Q) + O(1) \quad \text{by (6.1.2)}\\
&= h_{C\times C,\mathcal{O}(\{P_0\}\times C)}(P,Q) + h_{C\times C,\mathcal{O}(C\times\{P_0\})}(P,Q) - h_\Delta(P,Q) + O(1) \text{ by Proposition 5.1.3.(ii)}\\
&= h_{C,[P_0]}(P) + h_{C,[P_0]}(Q) - h_{C\times C,\Delta}(P,Q) + O(1) \text{ by Proposition 5.1.3.(iii)}^{[2]}\\
&= \frac{1}{g}h_{C,j^*\Theta^-}(P) + \frac{1}{g}h_{C,j^*\Theta^-}(Q) - h_{C\times C,\Delta}(P,Q) + O(1) \text{ by Lemma 6.1.2.(ii) and Proposition 5.1.3.(iii)}\\
&= \frac{1}{g}\hat{h}_{J,\Theta^-}(z) + \frac{1}{g}\hat{h}_{J,\Theta^-}(w) - h_{C\times C,\Delta}(P,Q) + O(1) \text{ by Proposition 5.1.3.(iii)}\\
&= \frac{1}{2g}\left(\hat{h}_{J,\Theta+\Theta^-}(z) - \hat{h}_{J,\Theta-\Theta^-}(z)\right) + \frac{1}{2g}\left(\hat{h}_{J,\Theta+\Theta^-}(w) - \hat{h}_{J,\Theta-\Theta^-}(w)\right) - h_{C\times C,\Delta}(P,Q) + O(1)\\
&\qquad\qquad \text{by Proposition 5.1.3.(ii)}\\
&= \frac{1}{2g}\left(|z|^2 - \hat{h}_{\Theta-\Theta^-}(z)\right) + \frac{1}{2g}\left(|w|^2 - \hat{h}_{\Theta-\Theta^-}(w)\right) - h_\Delta(P,Q) + O(1).
\end{aligned}$$

So

$$h_{C\times C,\Delta}(P,Q) = \frac{1}{2g}|j(P)|^2 + \frac{1}{2g}|j(Q)|^2 - \langle j(P), j(Q)\rangle - \frac{1}{2g}\hat{h}_{\Theta-\Theta^-}(j(P)) - \frac{1}{2g}\hat{h}_{\Theta-\Theta^-}(j(Q)) + O(1), \quad (6.1.4)$$

from which we can conclude. □

The equality (6.1.3) gives a strong restriction on the distribution of points on C. For the left hand side, we have[3] $h_{C\times C,\Delta}(P,Q) \geq 0$ for all $P \neq Q$, since Δ is an effective divisor on $C \times C$ and (P,Q) is not in $\mathrm{supp}(\Delta)$. For the right hand side, the main term is an *indefinite* quadratic form (since $g \geq 2$) and hence can *à priori* attain all negative values, especially when $j(P)$ and $j(Q)$ both have large norms and are "too close" to each other. The equality (6.1.3) then confirms that this last case cannot happen: If $|j(P)|$ and $|j(Q)|$ are both very large, then either the angle between $j(P)$ and $j(Q)$ is large, or $\max\{|j(P)|, |j(Q)|\}$ is significantly larger than $\min\{|j(P)|, |j(Q)|\}$.

6.1.2 Mumford's Gap Principle

Theorem 6.1.5 (Mumford's Inequality)**.** *Let $\epsilon > 0$. There exists a constant $R = R(C,\epsilon) > 0$ such that the following property holds true. Consider all pairs of distinct points $P, Q \in C(\overline{\mathbb{Q}})$ satisfying:*

(i) $|j(Q)| \geq |j(P)|$;

(ii) the angle between $j(P)$ and $j(Q)$ is at most $\cos^{-1}(3/4+\epsilon)$. In other words, $\langle j(P), j(Q)\rangle \geq (3/4+\epsilon)|j(P)||j(Q)|$.

If $|j(P)| \geq R$, then

$$|j(Q)| \geq 2|j(P)|. \quad (6.1.5)$$

Proof. By assumption, $P \neq Q$. So for a suitable representative height function, we have $h_{C\times C,\Delta}(P,Q) \geq 0$ since Δ is an effective divisor on $C \times C$ and (P,Q) is not in $\mathrm{supp}(\Delta)$.

[3] For a representative of the height function.

Denote by $z = j(P)$ and $w = j(Q)$. In addition to the last paragraph, by Proposition 6.1.4 and since $|w| \geq |z|$ (hypothesis (i)), we have

$$\langle z, w\rangle \leq \frac{1}{2g}|z|^2 + \frac{1}{2g}|w|^2 + O(|w| + 1).$$

Divide both sides by $|z||w|$. Hypothesis (ii) then implies

$$\frac{3}{4} + \epsilon \leq \frac{1}{2g}\left(\frac{|w|}{|z|} + \frac{|z|}{|w|}\right) + O\left(\frac{1}{|z|}\right) \leq \frac{1}{2g}\left(\frac{|w|}{|z|} + 1\right) + O\left(\frac{1}{|z|}\right),$$

from which we have

$$\frac{|w|}{|z|} \geq \frac{3}{2}g - 1 + 2g\epsilon - O\left(\frac{1}{|z|}\right) \geq 2 + 2g\epsilon - O\left(\frac{1}{|z|}\right).$$

So for $R \gg 1$, the hypothesis $|z| \geq R$ implies $|w| \geq 2|z|$. □

Mumford's Gap Principle allows us to continue the proof of the Mordell Conjecture as follows. In the Euclidean space $(J(K) \otimes \mathbb{R}, |\cdot| = \hat{h}^{1/2}_{\Theta+\Theta^-})$, the image of $J(K) \to J(K) \otimes \mathbb{R}$ is a lattice.[4] Consider the ball centered at the origin of radius R in the Euclidean space. We divide the points in $C(K)$ into:

- (small points) $\{P \in C(K) : \hat{h}_{\Theta+\Theta^-}(j(P)) < R\}$;
- (large points) $\{P \in C(K) : \hat{h}_{\Theta+\Theta^-}(j(P)) \geq R\}$.

Notice that the set of small points is immediately a finite set since it contains lattice points in a bounded ball. Now to prove the Mordell Conjecture, it suffices to prove the finiteness of the large points.

To do this, we divide $J(K) \otimes \mathbb{R}$ into cones according to the angles between distinct points given by hypothesis (ii) of Mumford's Gap Principle. More precisely, we divide $J(K) \otimes \mathbb{R}$ into $7^{\dim J(K)\otimes\mathbb{R}}$ cones such that the angle between each two points in a same cone is $\leq \cos^{-1}(3/4+\epsilon)$. Mumford's Gap Principle then says that in each cone, the norms of large points increase *rapidly*.

6.2 Vojta's Inequality: Statement and Consequence

Theorem 6.2.1 (Vojta's Inequality)**.** *Let $\epsilon > 0$. There exist two constants $R = R(C, \epsilon) > 0$ and $\kappa = \kappa(g, \epsilon) > 0$ such that the following property holds true. Consider all pairs of distinct points $P, Q \in C(\overline{\mathbb{Q}})$ satisfying:*

(i) $|j(Q)| \geq |j(P)|$;

(ii) the angle between $j(P)$ and $j(Q)$ is at most $\cos^{-1}(3/4+\epsilon)$*. In other words,* $\langle j(P), j(Q)\rangle \geq (3/4 + \epsilon)|j(P)||j(Q)|$.

If $|j(P)| \geq R$, then

$$|j(Q)| \leq \kappa|j(P)|. \tag{6.2.1}$$

[4] Two points in $J(K)$ have the same image in $J(K) \otimes \mathbb{R}$ if and only if they differ from a torsion point. Since $J(K)$ has only finitely many torsion points (by the Mordell–Weil Theorem), viewing $j(C(K))$ as a subset of $J(K)$ or as a subset of $J(K) \otimes \mathbb{R}$ does not change the finiteness.

The constant R here may be different from the one from Mumford's Gap Principle. However, we can replace them by the maximum and assume that it is the same constant.

Proof of Theorem 6.1.1 assuming Vojta's Inequality. By the discussion of the last section, it suffices to prove the finiteness of the large points in each cone.

Assume $P_0, \dots, P_n$ are such points of non-decreasing norms. Then Mumford's Gap Principle (6.1.5) implies

$$|j(P_n)| \geq 2|j(P_{n-1})| \geq \cdots \geq 2^n |j(P_0)|.$$

On the other hand, Vojta's Inequality says $|j(P_n)| \leq \kappa |j(P_0)|$. So $2^n \leq \kappa$. Hence in each cone, there are $\leq \log \kappa / \log 2 + 1$ large points. Hence we are done. □

6.3 Vojta divisors

6.3.1 Vojta divisors and the generalized Mumford's Formula

Let Δ be the diagonal of $C \times C$; it is a divisor on $C \times C$. Set

$$\Delta' = \Delta - \{P_0\} \times C - C \times \{P_0\}, \tag{6.3.1}$$

which has degree 0 on each fiber of the natural projections $C \times C \to C$.

Definition 6.3.1. *A Vojta divisor is of the form*

$$V(d_1, d_2, d) = d_1\{P_0\} \times C + d_2 C \times \{P_0\} + d\Delta'$$

for some $d_1, d_2, d \in \mathbb{Z}_{>0}$.

As for Proposition 6.1.4, one can prove the following *generalized Mumford's Formula.*

Proposition 6.3.2. *We have, for all* $P, Q \in C(\overline{\mathbb{Q}})$,

$$\begin{aligned} h_{V(d_1,d_2,d)}(P,Q) =& \frac{d_1}{2g}|j(P)|^2 + \frac{d_2}{2g}|j(Q)|^2 - d\langle j(P), j(Q)\rangle \\ &+ d_1 O(|j(P)|) + d_2 O(|j(Q)|) + (d_1 + d_2 + d)O(1). \end{aligned} \tag{6.3.2}$$

The main term of the right hand side of this formula is a quadratic form, which is indefinite if $d_1 d_2 < g^2 d^2$.

The proof of Vojta's Inequality is inspired by the proof of *Roth's Theorem.* Roughly speaking, one has to find an upper bound and a lower bound of the evaluation of a suitable function at certain points, and then conclue using these two (repelling) bounds with the parameters tending to infinity. In our case, this function a height $h_{V(d_1,d_2,d)}$ on $C \times C$, these points are the pairs (P, Q) outside the diagonal with P, Q in a same cone Γ constructed at the end of §6.1.2, and the parameters are d_1, d_2, d. In other words, one wishes to prove an upper bound and a lower bound for $h_{V(d_1,d_2,d)}(P, Q)$ for d_1, d_2, d large enough and for $(P, Q) \in \Gamma \times \Gamma$ outside the diagonal.

The ***upper bound*** is given by (6.3.2).

To prove the lower bound, we need to construct a *small* section of the line bundle $\mathcal{O}(V(d_1, d_2, d))$ on $C \times C$, "small" in an appropriate sense which will be made precise in later sections. Our tool to do this is by a suitable version of Siegel's Lemma.[5]

[5] How this is inspired by the proof of Roth's Theorem: We constructed an auxiliary polynomial of small height in the proof of Roth's Theorem. In §4.4.3, we explained that each polynomial can be seen as a section of a suitable line bundle on $(\mathbb{P}^1)^m$. Here we will construct a section of some line bundle $\mathcal{O}(V(d_1, d_2, d))$ on $C \times C$.

6.3.2 Sections of Vojta divisors

We start by writing a Vojta divisor $V = V(d_1, d_2, d)$ as the difference of two very ample divisors on $C \times C$.

Fix an integer $N \gg 1$ such that $N[P_0]$ is a very ample divisor on C. Then it gives rise to a closed immesion

$$\phi_{N[P_0]} \colon C \to \mathbb{P}^n_K$$

such that $\phi^*_{N[P_0]}\mathcal{O}(1) \simeq \mathcal{O}(N[P_0])$. Thus we get a closed immersion

$$\psi = \phi_{N[P_0]} \times \phi_{N[P_0]} \colon C \times C \to \mathbb{P}^n_K \times \mathbb{P}^n_K$$

and, for all pairs of positive integers (δ_1, δ_2), we have as line bundles on $C \times C$

$$\mathcal{O}(\delta_1 N\{P_0\} \times C + \delta_2 NC \times \{P_0\}) \simeq \psi^*\mathcal{O}(\delta_1, \delta_2). \tag{6.3.3}$$

Fix an integer $M \gg 1$ such that $B := M(\{P_0\} \times C + C \times \{P_0\}) - \Delta'$ is a very ample divisor on $C \times C$. Then it gives rise to a closed immesion

$$\phi_B \colon C \times C \to \mathbb{P}^m_K$$

such that $\phi^*_B\mathcal{O}(1) \simeq \mathcal{O}(B)$.

Suppose

$$\texttt{(V1)} \qquad \delta_1 := \frac{d_1 + Md}{N} \text{ and } \delta_2 := \frac{d_2 + Md}{N} \text{ are two integers.}$$

This is satisfied up to adding d_1 with an integer bounded by N (same for d_2). Now the Vojta divisor can be written as the difference of two very ample divisors

$$V = V(d_1, d_2, d) = (\delta_1 N\{P_0\} \times C + \delta_2 NC \times \{P_0\}) - dB. \tag{6.3.4}$$

Thus for all $P, Q \in C(K)$, we have

$$\begin{aligned}[K : \mathbb{Q}]h_V(P, Q) &= [K : \mathbb{Q}](\delta_1 h_{N[P_0]}(P) + \delta_2 h_{N[P_0]}(Q) - dh_B(P, Q)) \\ &= [K : \mathbb{Q}] \left(\delta_1 h(\phi_{N[P_0]}(P)) + \delta_2 h(\phi_{N[P_0]}(Q)) - dh(\phi_B(P, Q))\right).\end{aligned}$$

Denote by $\phi_{N[P_0]}(P) = \mathbf{x} = [x_0 : \cdots : x_n]$, $\phi_{N[P_0]}(Q) = \mathbf{x}' = [x'_0 : \cdots : x'_n]$ et $\phi_B(P, Q) = \mathbf{y} = [y_0 : \cdots : y_m]$. Then the right hand side of the equality above equals

$$\delta_1 \sum_{v \in M_K} \max_j \log \|x_j\|_v + \delta_2 \sum_{v \in M_K} \max_{j'} \log \|x'_{j'}\|_v - d \sum_{v \in M_K} \max_i \log \|y_i\|_v = \sum_{v \in M_K} \min_i \max_{j,j'} \log \left\| \frac{x_j^{\delta_1} x_{j'}^{\prime \delta_2}}{y_i^d} \right\|_v.$$

Hence we have

$$h_V(P, Q) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in M_K} \min_i \max_{j,j'} \log \left\| \frac{x_j^{\delta_1} x_{j'}^{\prime \delta_2}}{y_i^d} \right\|_v. \tag{6.3.5}$$

Suppose furthermore

$$\begin{aligned}\texttt{(V2)} \qquad & \psi^* \colon H^0(\mathbb{P}^n_K \times \mathbb{P}^n_K, \mathcal{O}(\delta_1, \delta_2)) \to H^0(C \times C, \psi^*\mathcal{O}(\delta_1, \delta_2)) \\ & \text{and } \phi^*_B \colon H^0(\mathbb{P}^m_K, \mathcal{O}(d)) \to H^0(C \times C, \mathcal{O}(dB)) \text{ are surjective.}\end{aligned}$$

This is satisfied for all $d_1, d_2, d \gg 1$.

Write $(\mathbf{x}, \mathbf{x}')$ for the coordinate of $\mathbb{P}^n_K \times \mathbb{P}^n_K$. Write $\mathbf{y}$ for the coordinate of $\mathbb{P}^m_K$. The following lemma is not hard to prove under `(V2)`.

Lemma 6.3.3. *For each global section $s \in H^0(C \times C, \mathcal{O}(V))$, there exist bi-homogeneous polynomials $F_i(\boldsymbol{x}, \boldsymbol{x}')$ with coefficients in K ($i \in \{0, \dots, m\}$) of bi-degree (δ_1, δ_2) such that*

$$s = \left.\frac{F_i(\boldsymbol{x}, \boldsymbol{x}')}{y_i^d}\right|_{C\times C} \qquad \text{for all } i \in \{0, \dots, m\}. \tag{6.3.6}$$

Conversely, if $F_i(\boldsymbol{x}, \boldsymbol{x}')$ ($i \in \{0, \dots, m\}$) are bi-homogeneous polynomials with coefficients in K of bi-degree (δ_1, δ_2) such that

$$(F_i(\boldsymbol{x}, \boldsymbol{x}')/y_i^d)|_{C\times C} = (F_j(\boldsymbol{x}, \boldsymbol{x}')/y_j^d)|_{C\times C} \tag{6.3.7}$$

for all i and j, then there exists a unique global section $s \in H^0(C \times C, \mathcal{O}(V))$ such that (6.3.6) *holds true.*

One importance of this lemma is that one can always take y_i^d as the denominator.

Proof. Let s be a global section of $\mathcal{O}(V)$. Then $s \otimes y_i^d|_{C\times C}$ is a global section of the line bundle $\mathcal{O}(V+dB) = \mathcal{O}(\delta_1 N\{P_0\}\times C + \delta_2 NC \times \{P_0\}) \simeq \psi^*\mathcal{O}(\delta_1, \delta_2)$; see (6.3.3) for the last isomorphism. Thus we obtain bi-homogeneous polynomials $F_i(\mathbf{x}, \mathbf{x}')$ of bi-degree (δ_1, δ_2) satisfying (6.3.6).

Conversely, assume $F_i(\mathbf{x}, \mathbf{x}')$ (with $i \in \{0, \dots, m\}$) are bi-homogeneous polynomials of bi-degree (δ_1, δ_2) satisfying (6.3.7). Then we get a meromorphic section s of $\mathcal{O}(V)$ by the formula (6.3.6). Notice that s does not depend on the choice of $i \in \{0, \dots, m\}$. Moreover, the poles of s are contained in the subvariety $\bigcap_{i=0}^m V(y_i) = \emptyset$. Thus s is a global section, and hence we are done. □

6.4 Lower bound: Construction of a small section

Assume

$$\texttt{(V3)} \qquad d_1 + d_2 > 4g - 4 \quad \text{and} \quad d_1 d_2 - gd^2 > \gamma d_1 d_2 \text{ for some } \gamma > 0.$$

Here γ depends only on d_1, d_2, d. We will make the choice later.

The goal of this section is to prove the following proposition, which constructs a small section.

Proposition 6.4.1. *Assume* (V1), (V2) *et* (V3). *Then there exists a global section $s \in H^0(C \times C, \mathcal{O}(V))$ associated with $\mathcal{F} = (F_0, \dots, F_m)$ such that*

$$h(\mathcal{F}) = O\left(\frac{d_1 + d_2}{\gamma}\right). \tag{6.4.1}$$

6.4.1 A coarse lower bound for $h_V(P, Q)$

We start with the following proposition, which gives a motivation for the desired construction of the small section.

Proposition 6.4.2. *Let $s \in H^0(C \times C, \mathcal{O}(V))$ be a global section which corresponds to $\mathcal{F} = (F_0, \dots, F_m)$ as in Lemme 6.3.3. Assume $s(P, Q) \neq 0$. Then*

$$h_V(P, Q) \geq -h(\mathcal{F}) - n \log\left((\delta_1 + n)(\delta_2 + n)\right), \tag{6.4.2}$$

where $h(\mathcal{F}) = \max_i h(F_i)$.

Proof. The Product Formula, since $s(P,Q) \neq 0$, implies

$$\sum_{v \in M_K} \log \|s(P,Q)\|_v = 0. \tag{6.4.3}$$

Denote by $\phi_{N[P_0]}(P) = \mathbf{x} = [x_0 : \cdots : x_n]$, $\phi_{N[P_0]}(Q) = \mathbf{x}' = [x'_0 : \cdots : x'_n]$ et $\phi_B(P,Q) = \mathbf{y} = [y_0 : \cdots : y_m]$. Then

$$\begin{aligned}
[K:\mathbb{Q}]h_V(P,Q) &= \sum_{v \in M_K} \min_i \max_{j,j'} \log \left\| \frac{x_j^{\delta_1} x_{j'}^{\prime \delta_2}}{y_i^d \cdot s(P,Q)} \right\|_v \quad \text{by (6.3.5) - (6.4.3)} \\
&= \sum_{v \in M_K} \min_i \max_{j,j'} \log \left\| \frac{x_j^{\delta_1} x_{j'}^{\prime \delta_2}}{F_i(x,x')} \right\|_v \quad \text{by (6.3.6)} \\
&= - \sum_{v \in M_K} \max_i \min_{j,j'} \log \left\| F_i \left(\left[\frac{x_0}{x_j}, \ldots, \frac{x_m}{x_j} \right], \left[\frac{x'_0}{x'_{j'}}, \ldots, \frac{x'_m}{x'_{j'}} \right] \right) \right\|_v .
\end{aligned}$$

Fix v and i, and let $j_{v,i}$ be such that $\|x_j/x_{j_{v,i}}\|_v \leq 1$ for all j and let $j'_{v,i}$ be such that $\|x'_{j'}/x'_{j'_{v,i}}\|_v \leq 1$ for all j'. Then we have

$$\min_{j,j'} \log \left\| F_i \left(\left[\frac{x_0}{x_j}, \ldots, \frac{x_m}{x_j} \right], \left[\frac{x'_0}{x'_{j'}}, \ldots, \frac{x'_m}{x'_{j'}} \right] \right) \right\|_v \leq \log \left\| F_i \left(\left[\frac{x_0}{x_{j_{v,i}}}, \ldots, \frac{x_m}{x_{j_{v,i}}} \right], \left[\frac{x'_0}{x'_{j'_{v,i}}}, \ldots, \frac{x'_m}{x'_{j'_{v,i}}} \right] \right) \right\|_v .$$

If $v \nmid \infty$, then the right hand side is $\leq \log \max \|\text{coefficients of } F_i\|_v$. If $v|\infty$, we should also take into account the number of monomials, which is $\leq \binom{\delta_1+n}{n}\binom{\delta_2+n}{n} \leq (\delta_1+n)^n(\delta_2+n)^n$. So

$$\begin{aligned}
[K:\mathbb{Q}]h_V(P,Q) &\geq - \sum_{v \in M_K} \max_i \log \max \|\text{coefficients of } F_i\|_v - \sum_{v|\infty} \log \|(\delta_1+n)^n(\delta_2+n)^n\|_v \\
&= -[K:\mathbb{Q}]h(\mathcal{F}) - [K:\mathbb{Q}]n \log\left((\delta_1+n)(\delta_2+n)\right).
\end{aligned}$$

Hence we are done. □

By this proposition, to obtain the desired lower bound il would suffice to choose a global section s which satisfies the following two properties: (i) s has *small height*, *i.e.* $h(\mathcal{F})$ from (6.4.2) is small; (ii) s does not vanish at (P,Q).

To get (i), we will use Siegel's Lemma. For (ii), one should use some results from zero estimates (Roth's Lemma in our case) to conclude that s does not vanish at (P,Q) to a large order (and also adjust appropriately the lower bound (6.4.2)). We finish (i) in this section, and finish (ii) in the next sections.

6.4.2 Construction of a small section

Now we apply Siegel's Lemma to construct a small section. We start by recalling the basic version of Siegel's Lemma, Lemma 2.1.1.

Lemma 6.4.3. *Let $A = (a_{ij})$ be an $M' \times N'$-matrix with entries in $\mathbb{Z}$. Set $B = \max_{i,j} |a_{ij}|$. If $N' > M'$, then $\mathrm{Ker}(A)$ contains a non-zero vector $\mathbf{x} = (x_1, \ldots, x_{N'}) \in \mathbb{Z}^{N'}$ such that*

$$\max_j |x_j| \leq (N'B)^{\frac{M'}{N'-M'}}.$$

The number N' is the number of variables. In our case, by Lemma 6.3.3 it is $(m+1)$-times of the dimension of the space of all $F(\mathbf{x},\mathbf{x}')|_{C\times C}$ of bi-degree (δ_1,δ_2). Hence

$$N' \approx (m+1)\dim H^0(C\times C, \psi^*\mathcal{O}(\delta_1,\delta_2)) \geq (m+1)(N\delta_1+1-g)(N\delta_2+1-g).$$

Here we have used the Riemann–Roch Theorem for C to prove this bound. Indeed by `(V2)`, we have

$$\dim H^0(C\times C, \psi^*\mathcal{O}(\delta_1,\delta_2)) = (N\delta_1+1-g)(N\delta_2+1-g). \tag{6.4.4}$$

If $\mathrm{rk}A = M'$, then the number $N'-M' = \dim \mathrm{Ker}(A)$. In our case, this number is

$$\dim H^0(C\times C, \mathcal{O}(V(d_1,d_2,d))) \tag{6.4.5}$$

which is $\geq d_1d_2 - gd^2 + (g-1)(d_1+d_2)$ if $d_1+d_2 > 4g-4$.[6]

If $\mathrm{rk}(A) < M'$, then one can do elementary eliminations for the linear system such that M' becomes $\mathrm{rk}(A)$. Notice that this elimination can be done in such a way that $h(A) = \log H(A)$ increases at most linearly.

A consequence of `(V3)` is

$$d = O(d_1+d_2). \tag{6.4.6}$$

Sketch of Proof of Proposition 6.4.1. In order to apply Siegel's Lemma, we should transform the conditions (6.3.7) into a linear system A whose variables are the coefficients of the F_i's. This is possible because the curve C (resp. the surface $C\times C$) is a subvariety of $\mathbb{P}^n_K$ (resp. of $\mathbb{P}^m_K$) by the closed immersion $\phi_{N[P_0]}$ (resp. by the closed immersion ϕ_B), and hence is defined as the zero locus of some polynomials. One can check that $h(A) := \log H(A) = O(d)$; for this we use the fact that in the expression (6.3.6) of s the denominator is y_i^d.

Now Siegel's Lemma implies

$$h(\mathcal{F}) \leq \frac{(m+1)\dim H^0(C\times C, \psi^*\mathcal{O}(\delta_1,\delta_2))}{\dim H^0(C\times C, \mathcal{O}(V(d_1,d_2,d)))}(O(d)+\log O(1)).$$

Therefore by (6.4.4), (6.4.5), `(V1)` and `(V3)`, we have

$$h(\mathcal{F}) \leq \frac{N^2\delta_1\delta_2 + O(d_1+d_2+d)}{\gamma d_1d_2 + O(d_1+d_2)}O(d). \tag{6.4.7}$$

Recall that $N\delta_i = d_i + Md$; see `(V1)`. Thus one can conclude by a direct computation and (6.4.6). □

More detailed proof of Proposition 6.4.1. To prove this proposition, we wish to translate (6.3.7) in to a linear system defined by a matrix A with entries in K, whose unknowns are the coefficients of the F_i's.

The morphism $\phi_{N[P_0]}\colon C\to\mathbb{P}^n_K$ sends C to a closed irreducible subvariety of $\mathbb{P}^n_K$, and $\phi_{N[P_0]}(C)$ is the zero locus of some linear forms over $\mathbb{P}^n_K$. Thus the morphism $\phi_B\colon C\times C\to\mathbb{P}^m_K$ can be expressed as

$$\phi_B = [p_0(\mathbf{x},\mathbf{x}') : \cdots : p_m(\mathbf{x},\mathbf{x}')].$$

with $(\mathbf{x},\mathbf{x}')$ the coordinates of $\mathbb{P}^n_K\times\mathbb{P}^n_K$. Here the p_i's are fixed polynomials.

Now (6.3.7) can be translated into

$$p_i^dF_0|_{C\times C} = p_0^dF_i|_{C\times C} \tag{6.4.8}$$

for all $i\in\{1,\dots,m\}$.

[6] This is a consequence of the Riemann–Roch Theorem for $C\times C$. We assume this estimate in this course.

By Theorem 1.3.4 (applied to $f_1 = \cdots = f_m = p_i$), we have $|h(p_i^d) - dh(p_i)| \le d\deg(p_i)\log 2$. As the p_i's are fixed, we have then $h(p_i^d) = O(d)$.

We need to handle the restriction to $C \times C$ in (6.4.8). To achieve this, we perform the following standard operation. There exists a morphism $\xi\colon C \to \mathbb{P}^2_K$ which is birational to its image.[7] If we write $\mathbf{z} = [z_0 : z_1 : z_2]$ the coordinates of $\mathbb{P}^2_K$, then $\xi(C)$ is defined by a homogeneous polynomial $f(z_0, z_1, z_2) = z_2^N + f_1(z_0, z_1)z_2^{N-1} + \cdots + f_N(z_0, z_1)$; notice that the degree is N because C is embedded into $\mathbb{P}^n$ via $N[P_0]$. Notice that $K(C) = K(\xi(C))$. Consider the set

$$K[\xi(C)] := \{G \in K[z_0 : z_1 : z_2] : G \text{ homogeneous and } \deg_{z_2} G < N\}.$$

Each element in $K(\xi(C))$ can then be written as the quotient of two polynomials in $K[\xi(C)]$, so can each element in $K(C)$.

Now, instead of the whole $H^0(C \times C, \psi^*\mathcal{O}(\delta_1, \delta_2))$, we work with its subspace given by

$$W = \{F_i'(\mathbf{z}, \mathbf{z}') : \deg_{\mathbf{z}} F_i' \le \delta_1,\ \deg_{\mathbf{z}'} F_i' \le \delta_2,\ \deg_{z_2} F_i' < N,\ \deg_{z_2'} F_i' < N\}^{m+1}.$$

Then

$$\begin{aligned} \dim W &= (m+1)\left(N\delta_1 - \frac{1}{2}N(N-3)\right)\left(N\delta_2 - \frac{1}{2}N(N-3)\right) \\ &= (m+1)N^2\left(\delta_1 - \frac{N-3}{2}\right)\left(\delta_2 - \frac{N-3}{2}\right). \end{aligned} \tag{6.4.9}$$

The solution space is then $H^0(C \times C, \mathcal{O}(V)) \cap W$. We wish to have a lower bound on its dimension. This can be obtained by

$$\dim H^0(C \times C, \mathcal{O}(V)) \cap W \ge \dim H^0(C \times C, \mathcal{O}(V)) - (\dim H^0(C \times C, \psi^*\mathcal{O}(\delta_1, \delta_2)) - \dim W).$$

Thus we obtain, by (6.4.5), (6.4.4) and (6.4.9),

$$\dim H^0(C \times C, \mathcal{O}(V)) \cap W \ge \gamma d_1 d_2 + O(d_1 + d_2). \tag{6.4.10}$$

Under the pull back of $\xi \times \xi\colon C \times C \to \mathbb{P}^2 \times \mathbb{P}^2$, (6.4.8) becomes a system of equations of the form $p_i^d(\mathbf{z}, \mathbf{z}')F_0'(\mathbf{z}, \mathbf{z}') = p_0^d(\mathbf{z}, \mathbf{z}')F_i'(\mathbf{z}, \mathbf{z}')$ (without further restrictions!). By comparing the coefficients to each monomial, we obtain a linear system A whose unknowns are the coefficients of the (F_i')'s and the coefficients are obtained by the p_i^d's. From the discussion above, we then have $h(A) = O(d)$.

Now Siegel's Lemma implies

$$h(\mathcal{F}) \le \frac{(m+1)\dim W}{\dim \Gamma(C \times C, \mathcal{O}(V)) \cap W}(O(d) + \log O(1)).$$

Hence by (6.4.9), (6.4.10), `(V1)` and `(V3)`, we have

$$h(\mathcal{F}) \le \frac{N^2\delta_1\delta_2 + O(d_1 + d_2 + d)}{\gamma d_1 d_2 + O(d_1 + d_2)}O(d). \tag{6.4.11}$$

Recall that $N\delta_i = d_i + Md$; see `(V1)`. Thus one can conclude by a direct computation and (6.4.6). □

6.4.3 Conclusion for Vojta's Inequality under the extra hypothesis

Assume that the s constructed by Proposition 6.4.1 satisfies $s(P, Q) \neq 0$. Then (6.4.2) and (6.4.1) yield

$$h_V(P, Q) \ge -O\left(\frac{d_1 + d_2}{\gamma}\right).$$

[7] This is true for any variety X of dimension r: there exists a morphism $X \to \mathbb{P}^{r+1}$ under which X is birational to its image.

With the upper bound $h_V(P,Q)$ given by (6.3.2), we then have

$$-O\left(\frac{d_1+d_2}{\gamma}\right) \leq \frac{d_1}{2g}|j(P)|^2 + \frac{d_2}{2g}|j(Q)|^2 - d\langle j(P), j(Q)\rangle + d_1 O(|j(P)|) + d_2 O(|j(Q)|) + O(d_1+d_2). \quad (6.4.12)$$

Now let us make the choice of d_1, d_2 and d. Set, for an integer $D \in \mathbb{Z}_{>0}$ and a real number $\gamma_0 \in (0,1)$,

$$d_1 = \sqrt{g+\gamma_0}\frac{D}{|j(P)|^2} + O(1), \quad d_2 = \sqrt{g+\gamma_0}\frac{D}{|j(Q)|^2} + O(1), \quad (6.4.13)$$
$$d = \frac{D}{|j(P)||j(Q)|} + O(1).$$

Let $D \to \infty$. Notice that $d_1 d_2 - g d^2 > \gamma d_1 d_2$ for $\gamma = \gamma_0/(g+\gamma_0) + o(1)$.

Dividing (6.4.12) by D, we get

$$\frac{\langle j(P), j(Q)\rangle}{|j(P)||j(Q)|} \leq \frac{\sqrt{g+\gamma_0}}{g} + O(\frac{1}{|j(P)|} + \frac{1}{|j(Q)|}).$$

The right hand side is $\leq 3/4$ when $|j(P)|, |j(Q)| \gg 1$ because $g \geq 2$; when this happens, then the angle between $j(P)$ and $j(Q)$ is $\geq \cos^{-1}(3/4)$, under the extra hypothesis that *there exists a small section s which does not vanish at (P,Q)*. In other words, *under this extra hypothesis*, there exists at most 1 large point in each cone constructed at the end of §6.1.2, a conclusion much stronger than Vojta's Inequality!

6.5 Lower bound: Admissible pair and conclusion

In practice, we cannot directly find such a global section $s \in H^0(C \times C, \mathcal{O}(V))$ which does not vanish at (P,Q). But we still can show that the small section s we find in Proposition 6.4.1 does not vanish to a high order at (P,Q) in an appropriate sense.

Definition 6.5.1. *Let $s \in H^0(C\times C, \mathcal{O}(V)) \setminus \{0\}$ and $(P,Q) \in (C\times C)(\overline{\mathbb{Q}})$. A pair $(i_1^*, i_2^*) \in \mathbb{N}^2$ is said to be **admissible** for s at (P,Q) if $\partial_{i_1^*}\partial'_{i_2^*} s(P,Q) \neq 0$ and*

$$\partial_{i_1}\partial'_{i_2} s(P,Q) = 0 \quad \text{for all } i_1 \leq i_1^*,\ i_2 \leq i_2^* \text{ and } (i_1,i_2) \neq (i_1^*, i_2^*).$$

Here $\partial_i = (1/i!)\partial^i$ and $\partial'_i = (1/i!)\partial'^i$, and ∂ (resp. ∂') is a non-zero vector in the tangent space of C at P (resp. at Q).

Zero estimate results, for example *Roth's Lemma*, allows to prove:

Proposition 6.5.2. *There exists a constant $c = c(C, \phi_{N[P_0]}, \phi_B) > 0$ with the following property. Assume $\varepsilon \in (0, 1/\sqrt{2}]$ and $(P,Q) \in (C\times C)(\overline{\mathbb{Q}})$, with $|j(P)|, |j(Q)| \gg 1$, satisfy*

$$\varepsilon^2 d_1 \geq d_2 \quad \text{et} \quad \min\{d_1|j(P)|^2, d_2|j(Q)|^2\} \geq \frac{c}{\gamma\varepsilon^2} d_1.$$

Then for the section s constructed in Proposition 6.4.1, there exists an admissible pair (i_1^*, i_2^*) for s at (P,Q) such that

$$\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2} \leq 4N\varepsilon.$$

Next one needs to generalize the lower bound (6.4.2) removing the hypothesis $s(P,Q) \neq 0$.

Proposition 6.5.3. *Let $s \in H^0(C \times C, \mathcal{O}(V))$ be a global section which corresponds to $\mathcal{F} = (F_0, \dots, F_m)$ as in Lemme 6.3.3. Let (i_1^*, i_2^*) be an admissible pair for s at (P, Q). Then*

$$h_V(P,Q) \geq -h(\mathcal{F}) - O(i_1^*|j(P)|^2 + i_2^*|j(Q)|^2 + i_1^* + i_2^*) - O(\delta_1 + \delta_2). \tag{6.5.1}$$

With these propositions, we can finish the proof of Vojta's Inequality (Theorem 6.2.1).

Proof of Theorem 6.2.1. Assume $P, Q \in C(\overline{\mathbb{Q}})$ satisfy the hypotheses (i) and (ii).The inequality (6.4.12) becomes, by replacing (6.4.2) by (6.5.1),

$$\begin{aligned} &- O\left(\frac{d_1+d_2}{\gamma}\right) - O(i_1^*|j(P)|^2 + i_2^*|j(Q)|^2 + i_1^* + i_2^*) - O(\delta_1+\delta_2) \qquad (6.5.2)\\ \leq &\frac{d_1}{2g}|j(P)|^2 + \frac{d_2}{2g}|j(Q)|^2 - d\langle j(P), j(Q)\rangle + d_1 O(|j(P)|) + d_2 O(|j(Q)|) + O(d_1+d_2). \end{aligned}$$

Choose d_1, d_2, d as in (6.4.13), and also γ_0 and γ. Recall that $O(\delta_1 + \delta_2) = O(d_1 + d_2)$ by `(V1)`. Then the inequality above becomes

$$\frac{\langle j(P), j(Q)\rangle}{|j(P)||j(Q)|} \leq \frac{\sqrt{g+\gamma_0}}{g} + O\left(\frac{1}{|j(P)|} + \frac{1}{|j(Q)|}\right) + O\left(\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2}\right) + O\left(\frac{i_1^*+i_2^*}{D}\right) \tag{6.5.3}$$

Since $|j(Q)| \geq |j(P)|$ by hypothesis (i) and $\langle j(P), j(Q)\rangle > (3/4)|j(P)||j(Q)|$ by hypothesis (ii), we have

$$\frac{3}{4} < \frac{\sqrt{g+\gamma_0}}{g} + O\left(\frac{1}{|j(P)|}\right) + O\left(\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2}\right) + O\left((\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2})\frac{1}{|j(P)|^2}\right). \tag{6.5.4}$$

Set $\varepsilon = |j(P)|/|j(Q)|$. Then $\varepsilon \in (0, 1/\sqrt{2}]$ by Mumford's Gap Principle (Theorem 6.1.5). We have $\varepsilon^2 d_1 \geq d_2$ by choice of d_1 and d_2. The condition $\min\{d_1|j(P)|^2, d_2|j(Q)|^2\} \geq \frac{c}{\gamma\varepsilon^2} d_1$ is satisfied if $|j(P)| \gg 1$. Hence we can apply Proposition 6.5.2 and obtain a lower bound $\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2} \leq 4N\varepsilon$. Thus (6.5.4) implies

$$\frac{3}{4} < \frac{\sqrt{g+\gamma_0}}{g} + 4N\varepsilon + O\left(\frac{1}{|j(P)|}\right) + O\left(12N\varepsilon\frac{1}{|j(P)|^2}\right).$$

This gives a positive lower bound, say ε_0, for ε when $|j(P)| \gg 1$. More precisely there exists $R = R(C) > 0$ such that $|j(P)| \geq R \Rightarrow \varepsilon > \varepsilon_0$. Notice that ε_0 depends only on g and γ_0. Hence we can fix a $\gamma_0 \in (0,1)$ and then ε_0 depends only on g.

Set $\kappa = 1/\varepsilon_0 \geq 1$ which depends only on g. If $|j(P)| \geq R$, then we have $|j(Q)| \leq \kappa|j(P)|$ by the previous paragraph. This is precisely Vojta's Inequality. □

6.6 Proof of Proposition 6.5.2

We give the outline of the proof of Proposition 6.5.2. In particular, we explain how Roth's Lemma is applied.

6.6.1 Basic setup

In order to apply Roth's Lemma, we need to project down C to $\mathbb{P}^1_K$. Such a finite morphism $C \to \mathbb{P}^1$ exists in general. Here let us be more precise and write down the projection. This is helpful for the computation afterwards.

There exists a morphism $\xi\colon C \to \mathbb{P}^2_K$ which is birational to its image. Moreover, by a change of coordinates, we may and do assume that ξ is precisely given by the natural projection $\mathbb{P}^n \to \mathbb{P}^2$, $\mathbf{x} \mapsto [x_0 : x_1 : x_2]$. Then $\xi(C)$ is defined by a homogeneous polynomial $f(x_0, x_1, x_2) = x_2^N + f_1(x_0, x_1)x_2^{N-1} + \cdots + f_N(x_0, x_1)$; notice that the degree is N because C is embedded into $\mathbb{P}^n$ via $N[P_0]$. Moreover, we may assume that none of the coordinates x_0, x_1, x_2 vanishes identically on C.

Consider the finite morphism $\pi\colon C \to \mathbb{P}^1_K$, $\mathbf{x} \mapsto [x_0 : x_1]$; it has degree N, meaning that for all but at most finitely many $[x_0 : x_1] \in \mathbb{P}^1(\overline{\mathbb{Q}})$, the inverse image under π has cardinality N.

With this projection π in hand, we wish to accomplish the following constructions. Recall that the small section $s \in H^0(C \times C, \mathcal{O}(V))$ which we constructed in Proposition 6.4.1; it is associated with $\mathcal{F} = (F_0, \ldots, F_m)$ with

$$s = \left.\frac{F_i(\mathbf{x}, \mathbf{x}')}{y_i^d}\right|_{C\times C} \qquad \text{for all } i \in \{0, \ldots, m\}.$$

We wish to find an admissible pair (i_1^*, i_2^*) for s at $(P, Q) \in (C \times C)(\overline{\mathbb{Q}})$. Under the morphism $\pi \times \pi\colon C \times C \to \mathbb{P}^1 \times \mathbb{P}^1$, we are able to push down s to a bi-homogeneous polynomial (which is a section is a suitable line bundle on $\mathbb{P}^1 \times \mathbb{P}^1$). Moreover, we can show that this bi-homogeneous polynomial shares similar properties as s. On the other hand, this bi-homogeneous polynomial becomes a polynomial in 2 variables when restricted to a suitable affine chart. In the end, we apply Roth's Lemma to this 2-variable polynomial to conclude.

6.6.2 Push down of s

We will omit all the proofs for this push down. Familiarity with the intersection theory of divisors will be helpful for this construction.

We start by pushing down $sy_i^d|_{C\times C} = F_i(\mathbf{x}, \mathbf{x}')|_{C\times C}$. The results are:

Lemma 6.6.1. *There is a bi-homogeneous polynomial $G_i \in K[x_0, x_1; x_0', x_1']$ of bi-degree $(N^2\delta_1, N^2\delta_2)$ such that $(\pi \times \pi)_*\mathrm{div}(F_i|_{C\times C}) = \mathrm{div}(G_i)$.*

Lemma 6.6.2. *$h(G_i) \le N^2 h(F_i) + O(\delta_1 + \delta_2)$.*

From these lemmas, we get the following result on the push down of s.

Lemma 6.6.3. *There exists a bi-homogeneous polynomial $E \in K[x_0, x_1; x_0', x_1']$ of bi-degree (Nd_1, Nd_2) such that*

(i) $(\pi \times \pi)_*\mathrm{div}(s) = \mathrm{div}(E)$;

(ii) if (j_1^*, j_2^*) is an admissible pair of E at $(\pi(P), \pi(Q))$, then there exists an admissible pair (i_1^*, i_2^*) of s at (P, Q) such that $i_1^* \le j_1^*$ and $i_2^* \le j_2^*$;

(iii) $h(E) \le N^2 h(\mathcal{F}) + O(d_1 + d_2 + d)$.

In order to perform the push down, it is already convenient to use the affine coordinates. Let us explain it in more details.

Since $\mathbb{P}^n \to \mathbb{P}^2$, $\mathbf{x} \mapsto [x_0 : x_1 : x_2]$, is well-defined on C and f vanishes at the point $[x_0(P) : x_1(P) : x_2(P)]$, we have either $x_0(P) \neq 0$ or $x_1(P) \neq 0$. Without loss of generality, we may assume $x_0(P) \neq 0$. Similarly we may assume $x_0(Q) \neq 0$ also holds true. On $\{\mathbf{x} \in C(\overline{\mathbb{Q}}) : x_0 \neq 0\}$, we use the affine coordinates $\xi_1, \ldots, \xi_n$, where $\xi_j := (x_j/x_0)|_C$ for $j \in \{1, \ldots, n\}$. Also write $(\boldsymbol{\xi}, \boldsymbol{\xi}')$ for the affine coordinates of $\mathbb{P}^n_K \times \mathbb{P}^n_K$.

Since $f(x_0, x_1, x_2)$ is monic in x_2, we have that ξ_2 is integral over the ring $K[\xi_1]$.

Consider the function field $K(C \times C)$, which is a finite extension of $K(\xi_1, \xi_1')$ of degree N^2. Then in the lemmas above, we can take

$$G_i(\xi_1, \xi_1') := \mathrm{Norm}_{K(C\times C)/K(\xi_1,\xi_1')} F_i(\boldsymbol{\xi}, \boldsymbol{\xi}')$$

and

$$E(\xi_1, \xi_1') := \mathrm{Norm}_{K(C\times C)/K(\xi_1,\xi_1')} \frac{F_i(\boldsymbol{\xi}, \boldsymbol{\xi}')}{(y_i/y_0)^d}.$$

Notice that this definition does not depend on i. Moreover, $\frac{F_i(\boldsymbol{\xi},\boldsymbol{\xi}')}{(y_i/y_0)^d}$ has no poles on the locus of $C \times C$ where $x_0 x_0' \neq 0$. So $E \in K[\xi_1, \xi_1']$.

6.6.3 Application of Roth's Lemma

Let us finish the proof of Proposition 6.5.2.

We wish to apply Roth's Lemma (Lemma 3.4.1) to $m = 2$, the polynomial E of degree Nd_1, Nd_2, the point $(\pi(P), \pi(Q))$, and $\sigma = \varepsilon^2$. Let us check the conditions.

Condition (i) is satisfied by the assumption $\varepsilon^2 d_1 \geq d_2$.

Let us check condition (ii). Recall that $h(\mathcal{F}) = O\left(\frac{d_1+d_2}{\gamma}\right)$. Thus by Lemma 6.6.3.(iii), we have

$$h(E) \leq N^2 O\left(\frac{d_1 + d_2}{\gamma}\right) + O(d_1 + d_2 + d).$$

But $d = O(d_1 + d_2)$ by (6.4.6) and $d_1 \geq d_2$. So

$$h(E) \leq N^2 O(d_1/\gamma). \tag{6.6.1}$$

On the other hand, we have

$$\min\{Nd_1 h(\pi(P)), Nd_2 h(\pi(Q))\} = N \min\{d_1 h_{N[P_0]}(P), d_2 h_{N[P_0]}(Q)\} + O(d_1)$$

because $h_{N[P_0]} = h \circ \pi + O(1)$ by Proposition 5.1.3.(iii). But

$$h_{N[P_0]}(P) = \frac{N}{g} h_{g[P_0]}(P) + O(1) = \frac{N}{2g} |j(P)|^2 + O(|j(P)| + 1),$$

where the last equality follows from Lemma 6.1.2.(ii). Hence we have

$$\min\{Nd_1 h(\pi(P)), Nd_2 h(\pi(Q))\} = \frac{N^2}{2g} \min\{d_1 |j(P)|^2, d_2 |j(Q)|^2\} + d_1 O(|j(P)|) + d_2 O(|j(Q)|) + O(d_1).$$

Thus assuming $|j(P)|, |j(Q)| \gg 1$, there exists a constant $c > 0$ such that

$$\min\{d_1 |j(P)|^2, d_2 |j(Q)|^2\} \geq \frac{c}{\gamma \varepsilon^2} d_1 \Rightarrow \min\{Nd_1 h(\pi(P)), Nd_2 h(\pi(Q))\} \geq \varepsilon^{-2}(h(E) + 8Nd_1). \tag{6.6.2}$$

Now, we can invoke Roth's Lemma and conclude that there exists an admissible pair (j_1^*, j_2^*) of E at $(\pi(P), \pi(Q))$ such that

$$\frac{j_1^*}{Nd_1} + \frac{j_2^*}{Nd_2} \leq 4\varepsilon.$$

Hence we are done by Lemma 6.6.3.(ii).

6.7 Proof of Proposition 6.5.3

We give an outline of the proof of Proposition 6.5.3. It is a consequence of Lemma 6.7.1 and Lemma 6.7.3.

6.7.1 First step

Recall the closed immersion $\phi_{N[P_0]}\colon C \to \mathbb{P}^n_K$. For $\mathbb{P}^n_K \times \mathbb{P}^n_K$, write $(\mathbf{x}, \mathbf{x}')$ for the coordinate. Denote by

$$\xi_{ij} := (x_i/x_j)|_C, \qquad \xi'_{ij} := (x'_i/x'_j)|_C. \tag{6.7.1}$$

They are well-defined non-zero rational functions on C.

For $v \in M_K$, let j_v be the index j for which $\|\xi_{j0}(P)\|_v$ is largest and similarly j'_v for $\|\xi'_{j0}(Q)\|_v$.

Lemma 6.7.1. *Let $s \in H^0(C \times C, \mathcal{O}(V))$ be a global section which corresponds to $\mathcal{F} = (F_0, \dots, F_m)$ as in Lemme 6.3.3. Let (i_1^*, i_2^*) be an admissible pair for s at (P, Q). Then*

$$\begin{aligned}[K:\mathbb{Q}]h_V(P,Q) \geq & -h(\mathcal{F}) - n\log\left((\delta_1+n)(\delta_2+n)\right) \\ & - \sum_{v\in M_K} \max_{\{i_\lambda\}} \left(\sum_\lambda \max_\nu \log \|\partial_{i_\lambda} \xi_{\nu j_v}(P)\|_v \right) \\ & - \sum_{v\in M_K} \max_{\{i'_\lambda\}} \left(\sum_\lambda \max_{\nu'} \log \|\partial'_{i'_\lambda} \xi'_{\nu' j'_v}(Q)\|_v \right) \\ & - (\delta_1 + \delta_2 + i_1^* + i_2^*),\end{aligned}$$

where $\{i_\lambda\}$ and $\{i'_\lambda\}$ run over all partitions of i_1^* and i_2^*, i.e. $\sum_\lambda i_\lambda = i_1^*$ and $\sum_\lambda i'_\lambda = i_2^*$.

Proof. Recall from (6.3.5) that

$$h_V(P,Q) = \frac{1}{[K:\mathbb{Q}]} \sum_{v\in M_K} \min_i \max_{j,j'} \log \left\| \frac{x_j^{\delta_1} x_{j'}'^{\delta_2}}{y_i^d} \right\|_v. \tag{6.7.2}$$

Because (i_1^*, i_2^*) is an admissible pair and by Leibniz rule, we have

$$\left(\frac{y_i^d}{x_j^{\delta_1} x_{j'}'^{\delta_2}} \partial_{i_1^*} \partial'_{i_2^*} s \right)(P,Q) = \partial_{i_1^*} \partial'_{i_2^*} \left(\frac{y_i^d}{x_j^{\delta_1} x_{j'}'^{\delta_2}} s \right)(P,Q).$$

Write $\boldsymbol{\xi}_j := (\xi_{0j}, \xi_{1j}, \dots, \xi_{nj})$ and $\boldsymbol{\xi}'_{j'} := (\xi'_{0j'}, \xi'_{1j'}, \dots, \xi'_{nj'})$, then the right-hand side equals $\partial_{i_1^*} \partial'_{i_2^*} F_i(\boldsymbol{\xi}_j, \boldsymbol{\xi}'_{j'})(P,Q)$. The Product Formula, since $(\partial_{i_1^*} \partial'_{i_2^*} s)(P,Q) \neq 0$, implies

$$\sum_{v\in M_K} \log \|(\partial_{i_1^*} \partial'_{i_2^*} s)(P,Q)\|_v = 0. \tag{6.7.3}$$

Therefore

$$\begin{aligned}[K:\mathbb{Q}]h_V(P,Q) &= \sum_{v\in M_K} \min_i \max_{j,j'} \log \left\| \frac{x_j^{\delta_1} x_{j'}'^{\delta_2}}{y_i^d (\partial_{i_1^*} \partial'_{i_2^*} s)(P,Q)} \right\|_v \qquad \text{by (6.7.2) - (6.7.3)} \\ &= -\sum_{v\in M_K} \max_i \min_{j,j'} \log \|\partial_{i_1^*} \partial'_{i_2^*} F_i(\boldsymbol{\xi}_j, \boldsymbol{\xi}'_{j'})(P,Q)\|_v.\end{aligned}$$

The number of monomials of F_i is bounded by $\binom{\delta_1+n}{n}\binom{\delta_2+n}{n} \le (\delta_1+n)^n(\delta_2+n)^n$. Thus

$$\begin{aligned}[K:\mathbb{Q}]h_V(P,Q) \ge & -h(\mathcal{F}) - n\log\left((\delta_1+n)(\delta_2+n)\right) \\ & - \sum_{v\in M_K} \min_j \max_{|\mathbf{l}|=\delta_1} \log \left\| \partial_{i_1^*} \boldsymbol{\xi}_j^{\mathbf{l}}(P) \right\|_v \\ & - \sum_{v\in M_K} \min_j \max_{|\mathbf{l}'|=\delta_2} \log \left\| \partial'_{i_2^*} \boldsymbol{\xi}_{j'}^{\boldsymbol{\eta}'}(Q) \right\|_v\end{aligned}$$

Here $\boldsymbol{\xi}_j^{\mathbf{l}} = \xi_{0j}^{l_0}\cdots\xi_{nj}^{l_n}$ and similarly for $\boldsymbol{\xi}_{j'}^{\boldsymbol{\eta}'}$. Since for each v we take the minimum with respect to j and j', we may take instead $j = j_v$ and $j' = j'_v$.

Let us consider $\log\|\partial_{i_1^*}\boldsymbol{\xi}_j^{\mathbf{l}}(P)\|_v$. Leibniz's rule implies

$$\partial_{i_1^*}\boldsymbol{\xi}_j^{\mathbf{l}} = \sum \prod_{\nu=0}^{n} \prod_{\mu=1}^{l_\nu} \partial_{i_{\mu\nu}} \xi_{\nu j},$$

where $\sum_{\mu\nu} i_{\mu\nu} = i_1^*$ and the sum above runs over all possible $\{i_{\mu\nu}\}$'s. Since the total number of pairs $\mu\nu$ equals δ_1, the number of possibilities for $i_{\mu\nu}$ equals $\binom{\delta_1+i_1^*-1}{i_1^*} \le 2^{\delta_1+i_1^*}$. We only need to consider the case where $j = j_v$. Since $\|\xi_{j_v0}(P)\|_v$ is the largest $\|\xi_{j0}(P)\|_v$, we have $\|\xi_{\nu j_v}(P)\|_v = \|\xi_{\nu 0}(P)/\xi_{j_v0}(P)\|_v \le 1$ for each ν. This allows us to get rid of the terms with $i_{\mu\nu} = 0$ in the estimates. Hence we get

$$\log\|\partial_{i_1^*}\boldsymbol{\xi}_j^{\mathbf{l}}(P)\|_v \le \max_{\{i_\lambda\}} \left(\sum_\lambda \max_\nu \log\|\partial_{i_\lambda}\xi_{\nu j_v}(P)\|_v \right) + (\delta_1 + i_1^*)\epsilon_v \log 2,$$

with $\epsilon_v = \begin{cases} [K_v:\mathbb{R}] & \text{if } v|\infty \\ 0 & \text{otherwise} \end{cases}$, and where $\{i_\lambda\}$ runs over all partitions of i_1^*. The same argument applies to $\log\|\partial'_{i_2^*}\boldsymbol{\xi}_{j'}^{\boldsymbol{\eta}'}(Q)\|_v$. Thus we can conclude. □

6.7.2 Local Eisenstein estimates

Next we wish bound from above the partial derivative evaluated at the given points.

Here is the idea of the local Eisenstein estimates. Recall that the partial derivative ∂ involves choosing a uniformizer x at $P \in C$. We may and do assume $x(P) = 0$ (up to replacing x by $x - x(P)$). Since $\mathrm{trdeg}K(C) = \mathrm{trdeg}K(x) = 1$, we have that $K(C)$ is a finite extension of $K(x)$. So any rational function $\xi \in K(C)$ is an algebraic function in x, and it satisfies a polynomial equation $p(x,\xi(x)) = 0$ for some $p \in K[x,t]$ with $\deg_t p \le [K(C):K(x)]$.

If $\frac{\partial p}{\partial t}(0,\xi(0)) \ne 0$, then the Implicit Function Theorem says that

$$\xi(x) = \sum_{k\ge 0} (\partial_k \xi(0)) x^i$$

in a neighborhood of 0, with $\partial_k = (1/k!)(\partial/\partial x)^k$. If we work over $\mathbb{C}$, then complex analysis says that the series above converges for all $|x| < \rho$, where

$$\rho = \liminf_{k\to\infty} |\partial_k \xi(0)|^{-1/k} > 0.$$

It follows that if we replace ρ by a smaller $\rho' > 0$, then we get a bound

$$|\partial_k \xi(0)| \le c\rho'^{-k} \qquad \text{for all } k \ge 0. \tag{6.7.4}$$

Here, the constant c takes care of the first few k's.

Theorem 6.7.2. *Let $p(x,t) \in K[x,t]$ be a polynomial in two variables with partial degrees $\leq d$. Let $\xi = \xi(x)$ be an algebraic function such that $p(x,\xi(x)) = 0$. Assume $\xi(0) \in K$, $\|\xi(0)\|_v \leq 1$ and $\frac{\partial p}{\partial t}(0,\xi(0)) \neq 0$. Then the Taylor coefficients of this algebraic function $\xi(x)$ satisfy*

$$\|\partial_k \xi(0)\|_v \leq \|8(d+1)^7\|_v^{k\eta_v} \left(\frac{\|p\|_v}{\|\frac{\partial p}{\partial t}(0,\xi(0))\|_v} \right)^{2k-1}$$

with $\eta_v := \begin{cases} 1 & \text{if } v|\infty \\ 0 & \text{otherwise} \end{cases}$.

6.7.3 Application of local Eisenstein estimates

Fix a non-constant rational function $f \in K(C)$. Since $\operatorname{trdeg}_K K(C) = \dim C = 1$, we have that $K(C)$ is a finite extension of $K(f)$. Each ξ_{ij} from (6.7.1) is thus algebraic over $K(f)$. Let $g_{ij}(x,t) \in K[x,t]$ be such that $g_{ij}(f,t) \in K(f)[t]$ is a minimal polynomial of ξ_{ij} over $K(f)$. In characteristic 0, every irreducible polynomial is separable, and hence $\frac{\partial}{\partial t} g_{ij}(f,\xi_{ij}) \neq 0$ in $K(C)$.

Let $Z \subseteq C(\overline{\mathbb{Q}})$ be the finite set consisting of:

(i) all zeros of x_j for $j \in \{0,\ldots,n\}$,

(ii) all poles of f,

(iii) the support of $\operatorname{div}(\mathrm{d}f)$,

(iv) the zeros of $\frac{\partial}{\partial t} g_{ij}(f,\xi_{ij})$.

Lemma 6.7.3. *If $P \notin Z$, then*

$$\sum_{v \in M_K} \max_{\{i_\lambda\}} \left(\sum_\lambda \max_\nu \log \|\partial_{i_\lambda} \xi_{\nu j_v}(P)\|_v \right) \ll i_1^*(|j(P)|^2 + 1),$$

with the constant in $\ll$ independent of P and i_1^*.

Proof. Denote by $\eta_v := \begin{cases} 1 & \text{if } v|\infty \\ 0 & \text{otherwise} \end{cases}$.

Let $P \notin Z$. Set $p_{ij}(x,t) := g_{ij}(x + f(P), t + \xi_{ij}(P))$. Then $p_{ij}(0,0) = 0$ and $\frac{\partial}{\partial t} p_{ij}(0,0) \neq 0$.

Apply the local Eisenstein theorem (Theorem 6.7.2) to ξ_{ij} and P. Then we get, for all $k \geq 1$,

$$\|\partial_k \xi_{ij}(P)\|_v \leq \|C_2\|_v^{k\eta_v} \left(\frac{\|p_{ij}\|_v}{\|\frac{\partial}{\partial t} p_{ij}(0,0)\|_v} \right)^{2k-1} \tag{6.7.5}$$

with C_2 depending only on $\deg g_{ij}$. Thus one gets

$$\begin{aligned}
&\sum_{v \in M_K} \max_{\{i_\lambda\}} \left(\sum_\lambda \max_\nu \log \|\partial_{i_\lambda} \xi_{\nu j_v}(P)\|_v \right) \\
\leq & i_1^* \log C_2 + \sum_{v \in M_K} \max_{\{i_\lambda\}} \sum_\lambda \max_\nu (2i_\lambda - 1) \left(\log^+ \|p_{\nu j_v}\|_v + \log^+ \left\| \frac{1}{\frac{\partial}{\partial t} p_{\nu j_v}(0,0)} \right\|_v \right) \\
\leq & i_1^* \log C_2 + 2i_1^* \sum_{v \in M_K} \sum_{i,j} \left(\log^+ \|p_{ij}\|_v + \log^+ \left\| \frac{1}{\frac{\partial}{\partial t} p_{ij}(0,0)} \right\|_v \right).
\end{aligned}$$

For $\|p_{ij}\|_v$, one gets easily the following upper bound by construction

$$\|p_{ij}\|_v \leq \|C_1\|_v^{\eta_v} \max\{1, \|f(P)\|_v, \|\xi_{ij}(P)\|_v\}^{\deg g_{ij}} \|g_{ij}\|_v \tag{6.7.6}$$

with $C_1 \leq (\deg g_{ij} + 1)^2 2^{\deg g_{ij}}$.

On the other hand,

$$\begin{aligned}
[K:\mathbb{Q}] \sum_{v\in M_K} \sum_{i,j} \log^+ \left\| \frac{1}{\frac{\partial}{\partial t} p_{ij}(0,0)} \right\|_v &= \sum_{i,j} h\left(\frac{1}{\frac{\partial}{\partial t} g_{ij}(f(P), \xi_{ij}(P))} \right) \\
&= \sum_{i,j} h\left(\frac{\partial}{\partial t} g_{ij}(f(P), \xi_{ij}(P)) \right) \\
&\leq C_3 \max\{h(f(P)), h(\xi_{ij}(P))\}
\end{aligned}$$

with C_3 depending only on $h(g_{ij})$ and $\deg g_{ij}$.

Putting these together, we obtain

$$\sum_{v\in M_K} \max_{\{i_\lambda\}} \left(\sum_\lambda \max_\nu \log \|\partial_{i_\lambda} \xi_{\nu j_v}(P)\|_v \right) \ll i_1^* \max\{h(f(P)), h(\xi_{ij}(P))\}. \tag{6.7.7}$$

It remains to show that $\max\{1, h(f(P)), h(\xi_{ij}(P))\} \ll |j(P)|^2 + 1$. For this, notice that (f, ξ_{ij}) can be viewed as the first two affine coordinates of a morphism $\varphi\colon C \to \mathbb{P}^r$ for some r. Hence $\max\{h(f(P)), h(\xi_{ij}(P))\} = h([1 : f(P) : \xi_{ij}(P)]) \leq h \circ \varphi(P) = h_{\varphi^*\mathcal{O}(1)}(P) + O(1)$. But since $N[P_0]$ is very ample in C, we have that $\mathcal{O}(kN[P_0]) \otimes \varphi^*\mathcal{O}(1)^\vee$ is very ample for some $k \gg 1$, and therefore $h_{\varphi^*\mathcal{O}(1)}(P) \leq k h_{N[P_0]}(P) + O(1)$. Hence

$$\max\{h(f(P)), h(\xi_{ij}(P))\} \ll h_{N[P_0]}(P) + 1.$$

Now the conclusion follows from Lemma 6.1.2.(ii), which implies $h_{N[P_0]}(P) = \frac{N}{2g}|j(P)|^2 + O(|j(P)| + 1)$. □