Heights

Lectures by Fabien Pazuki

Spring semester 2014

These are my notes from a course given by Fabien Pazuki at Université de Bordeaux. All typos and mistakes are mine, so please report them to cadadr@gmail.com. The last version of this text can be found at https://cadadr.org/

The recommended books:

- Marc Hindry, Joseph H. Silverman, *Diophantine Geometry: An Introduction*
- Enrico Bombieri, Walter Gubler, *Heights in Diophantine Geometry*

Contents

Heights is a fundamental tool in proving finiteness results in Diophantine geometry and counting the resulting finite sets.

First we would like to define the height $H(\alpha)$ of an algebraic number $\alpha \in \overline{\mathbb{Q}}$, in a way that there are finitely many numbers α with bounded height $H(\alpha) \le C$ and degree. For this one should consider the absolute values of α.

For example, the number $\frac{2014}{2013}$ is very close to 1 with respect to the usual absolute value $|\cdot|$. However, it carries a lot of arithmetic information, since it contains primes $3 \cdot 11 \cdot 61 = 2013$ and $2 \cdot 19 \cdot 53 = 2014$. So for instance its 2-adic or 3-adic absolute value is far from 1. The idea of a height is to put all the absolute values together.

1 Absolute values and product formula

First we recall some definitions and facts about absolute values on fields.

Definition 1.1. Let K be a field. An **absolute value** on K is a function $|\cdot| : K \to \mathbb{R}_{\ge 0}$ satisfying the following properties:

1. $|\alpha| = 0$ iff $\alpha = 0$.
2. Multiplicativity: $|\alpha\beta| = |\alpha| \cdot |\beta|$ for all $\alpha, \beta \in K$.
3. Triangle inequality: $|\alpha + \beta| \le |\alpha| + |\beta|$ for all $\alpha, \beta \in K$.

If holds a stronger inequality $|\alpha+\beta| \le \max\{|\alpha|, |\beta|\}$, then the absolute value is called **nonarchimedian**.

Definition 1.2. Two absolute values $|\cdot|_1$ and $|\cdot|_2$ on K are **equivalent** if there exists some number $\lambda > 0$ such that $|\cdot|_1 = |\cdot|_2^{\lambda}$.

An equivalence class of absolute values on K is called a **place**.

Theorem 1.3 (Ostrowski). *On $\mathbb{Q}$ any nontrivial absolute value is equivalent to one of the following:*

- *The usual archimedian absolute value, denoted $|\cdot|_\infty$.*
- *A p-adic absolute value $|\cdot|_p$ for some prime p.*

Recall that a **p-adic absolute value** is defined as follows: for a number $\frac{a}{b}$ we can write $\frac{a}{b} = p^{\ell}\frac{a'}{b'}$ where both a' and b' are not divisible by p, and then

$$\left|\frac{a}{b}\right|_p = \left|p^{\ell}\frac{a'}{b'}\right|_p := p^{-\ell}.$$

p-adic absolute values are nonarchimedian.

To simplify things we are going to work solely with number fields (even though much results are similar and even easier over global function fields). In these notes K always denotes a finite algebraic extension of $\mathbb{Q}$.

Proposition 1.4. *Let L/K be a finite extension of fields. Then an absolute value $|\cdot|_v$ on K can be extended to L by setting*

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

This actually allows us to extend $|\cdot|_v$ on $\overline{\mathbb{Q}}$, by taking L/K to be any number field containing x.

In general, over a place $|\cdot|_v$ on K there are several places $|\cdot|_w$ on L extending $|\cdot|_v$. We write "$w \mid v$" in this case.

Proposition 1.5. *Let L/K be a separable extension and let $|\cdot|_v$ be an absolute value on K. Then*

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

where by K_v we denote the completion of K with respect to $|\cdot|_v$.

For norms one has

$$\prod_{w|v} N_{L_w/K_v}(x) = N_{L/K}(x).$$

On the places $|\cdot|_w$ lying over $|\cdot|_v$ the Galois group $\mathrm{Gal}(L/K)$ acts by $|\cdot|_{\sigma\cdot w} := |\sigma(\cdot)|_w$. Namely, one has the following.

Proposition 1.6. *Let L/K be a finite Galois extension. Let $|\cdot|_{w_1}$ and $|\cdot|_{w_2}$ be two places extending a place $|\cdot|_v$ on K. Then there exists a unique element $\sigma \in \mathrm{Gal}(L/K)$ such that $|\cdot|_{w_2} = |\sigma(\cdot)|_{w_1}$.*

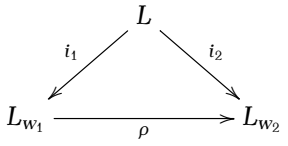
Proof. The places $|\cdot|_{w_{1,2}}$ correspond to embeddings $i_{1,2} \colon L \hookrightarrow L_{w_{1,2}}$ over K.

The extension L/K is separable. Let $L = K(\xi)$ for some $\xi \in L$ and let $f \in K[X]$ be the minimal polynomial of ξ over K. Consider the factorization of f over K_v:

$$f(X) = f_1(X) \cdots f_k(X) \in K_v[X].$$

Then the extensions L_{w_1}/K_v and L_{w_2}/K_v are of the form $K_v(\text{roots of } f_i)$. We have $L_{w_1} = L_{w_2}$, since L/K is Galois, so all roots of f are contained in $L_{w_{1,2}}$.

There is a unique element $\rho \in \mathrm{Gal}(L_{w_{1,2}}/K)$ such that $i_2 = \rho \circ i_1$.



$$i_1(\xi) \longmapsto i_2(\xi)$$

Then there exists a unique element $\sigma \in \mathrm{Gal}(L/K)$ such that $\rho \circ i_1 = i_1 \circ \sigma$, and

$$|x|_{w_2} = |i_2(x)|_v = |\rho \circ i_1(x)|_v = |i_1 \circ \sigma(x)|_v = |\sigma(x)|_{w_1}. \qquad \square$$

Definition 1.7. For a number field K an embedding $\sigma \colon K \hookrightarrow \mathbb{C}$ is said to be **real** if $\sigma(K) \subseteq \mathbb{R}$ and **complex** if $\sigma(K) \nsubseteq \mathbb{R}$. If σ is real, then it corresponds to a place $|\cdot|_v$ on K, and this $|\cdot|_v$ is said to be **real**.

If σ is complex, then the pair of σ and its complex conjugate $\overline{\sigma}$ corresponds to one place $|\cdot|_v$ on K. This $|\cdot|_v$ is said to be **complex**.

Observe that one has

$$[K_v : \mathbb{Q}_v] = \begin{cases} 1, & |\cdot|_v \text{ is real,} \\ 2, & |\cdot|_v \text{ is complex.} \end{cases}$$

Proposition 1.8 (Product formula)**.** *Let K be a number field. It is possible to choose a set M_K of representatives of equivalence classes of absolute values on K in such a way that for all $x \in K^\times$ holds*

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

where $d_v := [K_v : \mathbb{Q}_v]$.

Remark 1.9. Some authors, e.g. Bombieri and Gubler, normalize the absolute values by the local degrees d_v putting $\|x\|_v := |x|_v^{d_v}$, so that the product formula reads $\prod_v \|x\|_v = 1$.

For $\mathbb{Q}$ these normalized absolute values are the standard archimedian absolute value $|\cdot|_\infty$ and the standard p-adic absolute values $|\cdot|_p$ for all primes p. So for $\mathbb{Q}$ the statement is trivial: by multiplicativity, it is enough to check the formula for a prime q. It has absolute value 1 with respect to $|\cdot|_p$ for finite primes $p \neq q$; absolute value $1/q$ with respect to $|\cdot|_q$; and absolute value q with respect to $|\cdot|_\infty$. So the product is 1.

And for a field extension $K/\mathbb{Q}$ the standard way to extend absolute values

$$|x|_v := |N_{K_v/\mathbb{Q}_v}(x)|_p^{1/d_v} \quad \text{for } v \mid p,\ x \in K$$

still gives the product formula:

$$\prod_{v \in M_K} |x|_v^{d_v} = \prod_{p \in M_\mathbb{Q}} \prod_{v \mid p} |N_{K_v/\mathbb{Q}_p}(x)|_p = \prod_{p \in M_\mathbb{Q}} |N_{K/\mathbb{Q}}(x)|_p = 1,$$

by proposition 1.5 and the product formula on $\mathbb{Q}$.

We will denote by M_K the set of places on K, with representatives chosen in such a way that the product formula holds. We will write M_K^∞ for the archimedian (infinite) places, and M_K^0 for the nonarchimedian (finite) places.

2 Heights of numbers

Definition 2.1. For an algebraic number $\alpha \in \overline{\mathbb{Q}}$ let K be a finite extension of $\mathbb{Q}$ containing α. Put the **height** of α to be

$$H(\alpha) := \prod_{v \in M_K} \max\{1,\ |\alpha|_v\}^{d_v/d},$$

where $d_v := [K_v : \mathbb{Q}_v]$ and $d := [K : \mathbb{Q}]$.

Observe that in the product only finitely many terms are not equal to one.

Definition 2.2. The **logarithmic height** is given by

$$h(\alpha) := \log H(\alpha) = \sum_{v \in M_K} \frac{d_v}{d} \log^+ |\alpha|_v,$$

where $\log^+ |\alpha| := \log \max\{1, |\alpha|\}$.

Observe that $h(\alpha)$ (and hence $H(\alpha)$) does not depend on the choice of K. Indeed, if we take a bigger field L/K, then with respect to L the height is

$$\begin{aligned}
\frac{1}{[L:\mathbb{Q}]} \sum_{w \in M_L} [L_w : \mathbb{Q}_w] \cdot \log^+ |\alpha|_w &= \frac{1}{[L:K]\cdot[K:\mathbb{Q}]} \sum_{w \in M_L} [L_w : K_v] \cdot [K_v : \mathbb{Q}_v] \cdot \log^+ |\alpha|_w \\
&= \frac{1}{[L:K]\cdot[K:\mathbb{Q}]} \sum_{v \in M_K} [K_v : \mathbb{Q}_v] \cdot \log^+ |\alpha|_v \underbrace{\sum_{w \mid v} [L_w : K_v]}_{[L:K]} \\
&= \frac{1}{[K:\mathbb{Q}]} \sum_{v \in M_K} [K_v : \mathbb{Q}_v] \cdot \log^+ |\alpha|_v.
\end{aligned}$$

Example 2.3. Let $\frac{m}{n}$ be a rational number, with m and n coprime. Then

$$H\left(\frac{m}{n}\right) = \prod_{v\in M_{\mathbb{Q}}} \max\left\{1,\ \left|\frac{m}{n}\right|_v\right\} = \max\{|m|_\infty, |n|_\infty\}.$$

Indeed, for a finite prime p one has

$$\max\left\{1,\ \left|\frac{m}{n}\right|_p\right\} = \begin{cases} p^k, & p^k \mid n, \\ 1, & \text{otherwise}. \end{cases}$$

So we see that

$$\prod_{p\in M^0_{\mathbb{Q}}} \max\left\{1,\ \left|\frac{m}{n}\right|_v\right\} = |n|_\infty.$$

For $|\cdot|_\infty$ one has

$$\max\left\{1,\ \left|\frac{m}{n}\right|_\infty\right\} = \begin{cases} \left|\frac{m}{n}\right|_\infty, & |m|_\infty > |n|_\infty, \\ 1, & \text{otherwise}. \end{cases}$$

So $H(\frac{m}{n}) = \max\{|m|_\infty, |n|_\infty\}$.

Observe that $H(\alpha) \geq 1$ and $h(\alpha) \geq 0$, since in the definition we put $\max\{1,\cdot\}$ and $\log^+|\cdot|$ respectively. Obviously, $h(\alpha) = 0$ for $\alpha = 1$. Also, we see that $h(\alpha) = 0$ whenever α is a root of unity.

Theorem 2.4 (Kronecker). *For an algebraic number $\alpha \in \overline{\mathbb{Q}}$ one has $h(\alpha) = 0$ iff α is a root of unity.*

Proof. If α is an n-th root of unity, then for any $v \in M_K$

$$|\alpha|_v^n = |\alpha^n|_v = |1|_v = 1,$$

so $|\alpha|_v = 1$, implying $h(\alpha) = 0$.

Conversely, assume $h(\alpha) = 0$. Then for all $v \in M_K$ (including the archimedian v) one has $|\alpha|_v \leq 1$, in particular α is an algebraic integer. Let $d := [\mathbb{Q}(\alpha) : \mathbb{Q}]$ be the degree of α and let $\alpha_1, \ldots, \alpha_d$ be the conjugates of α.

Remark 2.5. Recall the so-called "Newton's theorem" which says that any symmetric polynomial $F(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_d]$ (invariant under variable permutation) can be expressed as a polynomial with coefficients in $\mathbb{Z}$ in the **elementary symmetric polynomials**

$$\begin{aligned} s_1(X_1, \ldots, X_d) &= \sum_{1\leq i\leq d} X_i, \\ s_2(X_1, \ldots, X_d) &= \sum_{1\leq i<j\leq d} X_i X_j, \\ s_3(X_1, \ldots, X_d) &= \sum_{1\leq i<j<k\leq d} X_i X_j X_k, \\ &\vdots \\ s_d(X_1, \ldots, X_d) &= X_1 \cdots X_d. \end{aligned}$$

Now if α is an algebraic integer, then for its minimal polynomial the Bézout identities give

$$(X-\alpha_1)\cdots(X-\alpha_d) = X^d - s_1(\alpha_1, \ldots, \alpha_d)\, X^{d-1} + s_2(\alpha_1, \ldots, \alpha_d)\, X^{d-2} - \cdots + (-1)^d\, s_d(\alpha_1, \ldots, \alpha_d).$$

Where $\alpha_1, \ldots, \alpha_d$ are the conjugates of α. Since α is an algebraic integer, this means that the minimal polynomial of α has coefficients in $\mathbb{Z}$, thus $s_i(\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}$, and (by the Newton's theorem) the value of any symmetric polynomial at $\alpha_1, \ldots, \alpha_d$ is in $\mathbb{Z}$.

We consider the values of symmetric polynomials $s_i(\alpha_1^m, \dots, \alpha_d^m)$ for $m = 1, 2, 3, \dots$ Since $|\alpha_i^m|_v \le 1$, we estimate using the triangle inequality

$$\sum_{1 \le i \le d} |s_i(\alpha_1^m, \dots, \alpha_d^m)|_v \le \sum_{1 \le i \le d} \binom{d}{i} \le 2^d,$$

for both nonarchimedian and archimedian places.

Now since $s_i(\alpha_1^m, \dots, \alpha_d^m)$ is an integer, the bound above tells that there are finitely many possible values of $s_i(\alpha_1^m, \dots, \alpha_d^m)$ for a fixed degree of α, so finitely many possible minimal polynomials, and the sequence

$$1, \alpha, \alpha^2, \alpha^3, \dots$$

consists of finitely many numbers. Thus $\alpha^n = \alpha^m$ for some $m > n$, and either $\alpha = 0$, or $\alpha^{m-n} = 1$ and α is a root of unity. □

Remark 2.6. Note that if instead of $h(\alpha) = 0$ we assume $h(\alpha) \le C$, then we can also bound the coefficients of the minimal polynomial f_α in terms of C and $\deg \alpha$. We will come back to this in § 5.

Here are some basic properties of heights:

1. $h(\alpha\beta) \le h(\alpha) + h(\beta)$.

   This is because $|\alpha\beta|_v = |\alpha|_v \cdot |\beta|_v \le \max\{1, |\alpha|_v\} \cdot \max\{1, |\beta|_v\}$, and thus

   $$\log^+ |\alpha\beta|_v \le \log^+ |\alpha|_v + \log^+ |\beta|_v.$$

2. $h(\alpha_1 + \cdots + \alpha_r) \le h(\alpha_1) + \cdots + h(\alpha_r) + \log r$.

   This is proved similarly, but we note that for an archimedian absolute value we must use the triangle inequality, hence

   $$|\alpha_1 + \cdots + \alpha_r|_v \le r \cdot \max\{|\alpha_1|_v, \dots, |\alpha_n|_v\} \le r \cdot \max\{1, |\alpha_1|_v\} \cdots \max\{1, |\alpha_m|_v\},$$

   and so

   $$\log^+ |\alpha_1 + \cdots + \alpha_r|_v \le \log^+ |\alpha_1|_v + \cdots + \log^+ |\alpha_r|_v + \log r.$$

   For nonarchimedian absolute values there is no $\log r$ term. Thus

   $$h(\alpha_1 + \cdots + \alpha_r) \le \frac{1}{d} \sum_{v \in M_K^0} d_v (\log^+ |\alpha_1|_v + \cdots + \log^+ |\alpha_r|_v) + \frac{1}{d} \sum_{v \in M_K^\infty} d_v (\log^+ |\alpha_1|_v + \cdots + \log^+ |\alpha_r|_v + \log r).$$

   But $\sum_{v \in M_K^\infty} d_v = d$, so we have exactly $h(\alpha_1 + \cdots + \alpha_r) \le h(\alpha_1) + \cdots + h(\alpha_r) + \log r$. We will encounter often this situation when a bound has an extra constant, coming from the use of triangle inequality for archimedian places.

3. $h(\alpha^r) = |r| \cdot h(\alpha)$ for all $r \in \mathbb{Q}$ and $\alpha \in \overline{\mathbb{Q}}$.

   This is clear for $r > 0$, since in this case $\max\{1, |\alpha^r|_v\} = \max\{1, |\alpha|_v\}^r$. We need to show this property only for $r = -1$. Observe that $\log^+ |\alpha^{-1}|_v = -\log^- |\alpha|_v$, where we denote $\log^- |x| := \min\{0, \log |x|\}$. Then $\log |x| = \log^+ |x| + \log^- |x|$, and by the product formula

   $$h(\alpha) - h(\alpha^{-1}) = \frac{1}{d} \sum_{v \in M_K} d_v \log^+ |\alpha|_v + \frac{1}{d} \sum_{v \in M_K} d_v \log^- |\alpha|_v = \frac{1}{d} \sum_{v \in M_K} d_v \log |\alpha|_v = 0.$$

4. For all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ one has $h(\sigma(\alpha)) = h(\alpha)$.

   This is because the Galois group acts on places M_K by permutation (proposition 1.6).

Example 2.7. Let us compute the logarithmic height of $1/2$, treating it as a number in the field $K = \mathbb{Q}(i)$.

$$h(1/2) = \sum_{v \in M_{\mathbb{Q}(i)}} \frac{d_v}{d} \log^+ |1/2|_v.$$

Remark 2.8. Recall some facts about the Gaussian integers.

- The norm is given by $N_{\mathbb{Q}(i)/\mathbb{Q}}(a+bi) = a^2 + b^2$.
- In $\mathbb{Z}[i]$ there are four units ± 1, $\pm i$, and the following are the prime ideals:
  - $(1+i)$
  - $(a+bi)$ with $a^2 + b^2 \equiv p \equiv 1 \pmod 4$, where p is prime in $\mathbb{Z}$.
  - (p), where p is prime in $\mathbb{Z}$ and $p \equiv 3 \pmod 4$.
- In particular, $1+i$ and $1-i$ are primes in $\mathbb{Z}[i]$ lying over the prime 2 in $\mathbb{Z}$. Each $1 \pm i$ generate the same ideal, since they differ by a unit i.
  We have a ramification
  $$2\,\mathbb{Z}[i] = (1+i)^2\,\mathbb{Z}[i].$$
- $\mathbb{Q}(i)$ has two pairs of complex embeddings, giving only one archimedian absolute value
  $$|a+bi| = \sqrt{a^2+b^2}.$$

One gets

$$h(1/2) = \frac{1}{2} \sum_{v \in M_{\mathbb{Q}(i)}} d_v \log^+ |1/2|_v.$$

The absolute value $|1/2|_v$ can be distinct from 1 only for v lying over 2 or ∞. Over ∞ lies one absolute value with local degree 2, and over 2 lies one absolute value with local degree 2:

$$h(1/2) = \frac{1}{2}\left(2 \cdot \log^+ |1/2|_\infty + 2 \cdot \log^+ |1/2|_{1\pm i}\right).$$

Now $|1/2|_\infty < 1$, so $\log^+ |1/2|_\infty = 0$. The remaining term is

$$\log^+ |1/2|_{1\pm i} = \log^+ |2|_{1\pm i}^{-1} = \log^+ |N_{\mathbb{Q}(i)/\mathbb{Q}}(2)|_2^{-1/[\mathbb{Q}(i):\mathbb{Q}]} = \log^+ |4|_2^{-1/2} = \log 2.$$

So we conclude that $h(1/2) = \log 2$ (which is immediate if we treat $1/2$ as an element of $\mathbb{Q}$ and not of $\mathbb{Q}(i)$).

Example 2.9. Similarly, we can compute

$$h(1+i) = \frac{1}{2}\left(2 \cdot \log^+ |1+i|_\infty + 2 \cdot \underbrace{\log^+ |1+i|_{1\pm i}}_{=0}\right) = \log \sqrt{2}.$$

And also $h((1-i)^{-1}) = h(1-i) = \log \sqrt{2}$.

On the other hand,

$$h\left(\frac{1+i}{1-i}\right) = h(i) = 0,$$

because i is a root of unity in $\mathbb{Q}(i)$.

This example shows that $h(\alpha\beta)$ can be strictly less than $h(\alpha)+h(\beta)$, because we can have $\log^+(ab) < \log^+(a) + \log^+(b)$.

$$0 = h((1+i)(1-i)^{-1}) < h(1+i) + h((1-i)^{-1}) = \log 2.$$

3 Heights of polynomials

Definition 3.1. For a polynomial

$$f(T_1,\dots,T_n) = \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n}\, T_1^{i_1}\cdots T_n^{i_n} =: \sum_{\underline{i}} a_{\underline{i}}\, T^{\underline{i}} \in K[T_1,\dots,T_n]$$

we define its **height** to be

$$h(f) := \sum_{v\in M_K} \frac{d_v}{d} \log |f|_v,$$

where

$$|f|_v := \max_{\underline{i}} |a_{\underline{i}}|_v.$$

The following is immediate from the definition.

Proposition 3.2. *If f and g are polynomials in independent variables, then $h(fg) = h(f) + h(g)$.*

This is wrong in general. For instance,

$$h(X+1) = 0, \quad h((X+1)^2) = h(X^2+2X+1) = \log 2.$$

The problem is that $|fg|_v \neq |f|_v \cdot |g|_v$ for archimedian places. However, for nonarchimedian places this is true.

Proposition 3.3 (Gauss' lemma)**.** *Let v be a nonarchimedian place. Then $|fg|_v = |f|_v \cdot |g|_v$.*

Proof. By definition $|fg|_v \le |f|_v \cdot |g|_v$. We may assume that $|f|_v = |g|_v = 1$. For the sake of contradiction suppose $|fg|_v < 1$.
Assume f and g are polynomials in one variable. We look at the coefficients of fg:

$$f = \sum a_k T^k, \quad g = \sum b_\ell T^\ell, \quad fg = \sum c_j T^j,\ c_j = \sum_{k+\ell=j} a_k\, b_\ell.$$

Consider the smallest index j such that $|a_j|_v = 1$. Since $|c_j|_v < 1$ and $|a_k|_v < 1$ for any $k < j$, we have $|b_0|_v < 1$. Now we consider coefficients $c_{j+\ell} = \sum_{m+n=j+\ell} a_m\, b_n$, and by induction we conclude that $|b_\ell|_v < 1$, and so $|g|_v < 1$, which is a contradiction.

For the multivariate case $f(T_1,\dots,T_n)$, $g(T_1',\dots,T_m')$, let $d = \deg(fg)+1$. Consider polynomials

$$f(T, T^d, T^{d^2},\dots,T^{d^{n-1}}) \quad \text{and} \quad f(T, T^d, T^{d^2},\dots,T^{d^{m-1}}).$$

By the choice of d, there is no cancellation in terms, and we can apply the one variable case. □

Of course the interesting question is how the height $h(f_\alpha)$ of the minimal polynomial of an algebraic number is related to the height $h(\alpha)$. Unfortunately, one can show only an inequality $h(\alpha) \le h(f_\alpha) + C$.

Example 3.4. Let $\phi_n(X)$ be the n-th cyclotomic polynomial which has as its roots the primitive n-th roots of unity. It is an irreducible polynomial with integer coefficients, the minimal polynomial of any primitive n-th root of unity ζ_n. We have $h(\zeta_n) = 0$, but $h(\phi_n) = 0$ would imply that ϕ_n has only coefficients 0 or 1. If we examine the cyclotomic polynomials, they indeed seem to have coefficients 0 or 1:

$$\phi_2(X) = X+1, \quad \phi_3(X) = X^2+X+1, \quad \phi_4(X) = X^2+1,$$
$$\phi_5(X) = X^4+X^3+X^2+X+1, \quad \phi_6(X) = X^2-X+1, \quad \phi_7(X) = X^6+X^5+X^4+X^3+X^2+X+1,$$
$$\phi_8(X) = X^4+1, \quad \phi_9(X) = X^6+X^3+1, \quad \phi_{10}(X) = X^4-X^3+X^2-X+1,$$
$$\cdots\cdots\cdots$$

But it is not true starting from $n = 105$: the cyclotomic polynomial $\phi_{105}(X)$ has 2 among its coefficients:

$$\begin{aligned}\phi_{105}(X) =& X^{48} + X^{47} + X^{46} - X^{43} - X^{42} - 2X^{41} - X^{40} - X^{39} + X^{36} + X^{35} + X^{34} + X^{33} + X^{32} \\ &+ X^{31} - X^{28} - X^{26} - X^{24} - X^{22} - X^{20} + X^{17} + X^{16} + X^{15} + X^{14} + X^{13} + X^{12} \\ &- X^9 - X^8 - 2X^7 - X^6 - X^5 + X^2 + X + 1.\end{aligned}$$

Proposition 3.5. *Let $f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in K[X]$ be a polynomial and $|\cdot|_v$ be an absolute value on K. Set $|f|_v := \max\{|a_0|_v, \ldots, |a_n|_v\}$. Let α be a root of $f(X)$. Then*

$$|\alpha|_v \le \begin{cases} |f|_v, & v \text{ nonarchimedian,} \\ 2\,|f|_v, & v \text{ archimedian.} \end{cases}$$

Proof. Since $|f|_v$ is by definition the maximum of $|a_i|_v$, we have $|f|_v \ge 1$. If $|\alpha|_v < 1$, then

$$|\alpha|_v < |f|_v \le 2\,|f|_v,$$

and we are done.

Now for $|\alpha|_v \ge 1$ we consider the expression

$$\alpha^n = -\sum_{0 \le i \le n-1} a_i\,\alpha^i.$$

We take the absolute values $|\cdot|_v$ and estimate the right hand side. In the nonarchimedian case

$$|\alpha|_v^n = \left|\sum_{0 \le i \le n-1} a_i\,\alpha^i\right|_v \le |f|_v \cdot |\alpha|_v^{n-1},$$

thus $|\alpha|_v \le |f|_v$. (In the bound we indeed used that $|\alpha|_v \ge 1$.)

In the archimedian case we do the same estimates, but we have to use the triangle inequality. Observe that we can assume $|\alpha|_v > 2$, otherwise the claimed inequality is trivially true.

$$\begin{aligned}|\alpha|_v^n = \left|\sum_{0 \le i \le n-1} a_i\,\alpha^i\right|_v &\le \sum_{0 \le i \le n-1} |a_i|_v \cdot |\alpha|_v^i \\ &= |\alpha^{n-1}|_v \sum_{0 \le i \le n-1} |a_i|_v \cdot |\alpha|_v^{i-(n-1)} \\ &\le |\alpha^{n-1}|_v \cdot |f|_v \cdot \left(1 + \frac{1}{|\alpha|_v} + \frac{1}{|\alpha|_v^2} + \cdots + \frac{1}{|\alpha|_v^{n-1}}\right) \\ &\le |\alpha^{n-1}|_v \cdot |f|_v \cdot \left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots\right) \\ &\le 2 \cdot |\alpha^{n-1}|_v \cdot |f|_v.\end{aligned}$$

(Note the interesting trick; a simple-minded application of the triangle inequality gives immediately $|\alpha|_v \le n\,|f|_v$, but we were able to replace "n" with "2".) □

From the last bound the following follows immediately.

Proposition 3.6. $h(\alpha) \le h(f_\alpha) + \log 2$.

4 Mahler measure

We defined $h(f)$ for a polynomial, but with this definition $h(fg) \neq h(f) + h(g)$, because $|f\,g|_v \neq |f|_v \cdot |g|_v$ for archimedian places. To solve this issue, one can assign to f another quantity, which turns out to be more natural.

Definition 4.1. Let $f \in K[X_1, \dots, X_n]$ be a polynomial. Then its **Mahler measure** is given by

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

where $\mathbb{T} := \{e^{i\theta} \mid 0 \leq \theta < 2\pi\}$ and $d\mu := \frac{1}{2\pi}\, d\theta$.

It is now trivial from the definition (by linearity of integration) that $M(fg) = M(f) \cdot M(g)$.
Now for a number $\alpha \in \mathbb{C}$ we calculate

- If $|\alpha| \geq 1$, then
$$\int_{\mathbb{T}} \log |e^{i\theta} - \alpha|\, d\mu = \log |\alpha|,$$
and so $M(T - \alpha) = |\alpha|$.
- If $|\alpha| < 1$, then
$$M(T - \alpha) = M(T\,(1 - \alpha\, T^{-1})) = M(T)\, M(1 - \alpha\, T^{-1}) = 1.$$

So we get

$$M(T - \alpha) = \exp(\log^+ |\alpha|) = \begin{cases} |\alpha|, & |\alpha| \geq 1, \\ 1, & |\alpha| < 1. \end{cases} \tag{1}$$

If a polynomial factors as

$$f(T) = a_d\, (T - \alpha_1) \cdots (T - \alpha_d),$$

then we have for its Mahler measure

$$M(f) = M(a_d)\, M(T - \alpha_1) \cdots M(T - \alpha_d).$$

Taking logarithms and using (1), one gets the so-called **Jensen's formula**

$$\log M(f) = \log |a_d| + \log^+ |\alpha_1| + \cdots + \log^+ |\alpha_d|. \tag{2}$$

From this we obtain

$$M(f) = |a_d| \cdot \prod_{|\alpha_i| \geq 1} |\alpha_i|.$$

Sometimes this is taken as the definition of M.

We are interested in Mahler measure because of the following.

Proposition 4.2. *For an algebraic number $\alpha \in \overline{\mathbb{Q}}$ let $f_\alpha \in \mathbb{Z}[T]$ be its minimal polynomial. Then*

$$\log M(f_\alpha) = \deg \alpha \cdot h(\alpha).$$

This is a remarkable fact, although one can guess that something like this holds by looking at the Jensen's formula (2).

In particular, we have

$$M(f_\alpha) = \prod_{|\alpha_i| \geq 1} |\alpha_i| \geq \prod |\alpha_i| = |N_{\mathbb{Q}(\alpha)/\mathbb{Q}}(\alpha)|,$$

so we get a lower bound on the height

$$\log |N_{\mathbb{Q}(\alpha)/\mathbb{Q}}(\alpha)| \leq \log M(f_\alpha) = \deg \alpha \cdot h(\alpha). \tag{3}$$

Proof. Let

$$f_\alpha(T) = a_d\, T^d + \cdots + a_1\, T + a_0.$$

We want to use proposition 1.6 about the action of the Galois group on places. So let $K/\mathbb{Q}$ be a *Galois extension* containing α, so that $(\sigma a)_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})}$ contains every conjugate of α exactly $\frac{[K:\mathbb{Q}]}{d}$ times.

$$f_\alpha(T) = a_d\,(T-\alpha_1)\cdots(T-\alpha_d).$$

Taking absolute values, we get

$$|f_\alpha|_v = |a_d|_v \cdot |T-\alpha_1|_v \cdots |T-\alpha_d|_v.$$

Recall that $|f_\alpha|_v := \max_i |a_i|_v$. For a nonarchimedian place v we must have $|f_\alpha|_v = 1$, otherwise all coefficients are divisible by some p, contradicting the minimality of the polynomial. Further for a a nonarchimedian place we can apply Gauss' lemma which gives

$$1 = |f_\alpha|_v = |a_d|_v \cdot \prod_{1\le i\le d} \max\{1, |\alpha_i|_v\} = |a_d|_v \cdot \prod_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})} \max\{1, |\sigma\alpha|_v\}^{\frac{d}{[K:\mathbb{Q}]}}.$$

Taking logarithms and multiplying all by the local degree d_v, we get

$$0 = \log |a_d|_v^{d_v} + \sum_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})} \frac{d}{[K:\mathbb{Q}]} \log^+ |\sigma\alpha|_v^{d_v}. \tag{4}$$

Note that we also have (keeping in mind that $\mathrm{Gal}(K/\mathbb{Q})$ just permutes the places)

$$[K:\mathbb{Q}]\cdot h(\alpha) = \sum_{v\in M_K} \sum_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})} \log^+ |\sigma\alpha|_v^{d_v} = \sum_{v\in M_K^\infty} \sum_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})} \log^+ |\sigma\alpha|_v^{d_v} + \sum_{v\in M_K^0} \sum_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})} \log^+ |\sigma\alpha|_v^{d_v}.$$

Now using (4), we get

$$[K:\mathbb{Q}]\cdot h(\alpha) = \sum_{v\in M_K^\infty} \sum_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})} \log^+ |\sigma\alpha|_v^{d_v} - \sum_{v\in M_K^0} \frac{[K:\mathbb{Q}]}{d} \log |a_d|_v^{d_v}.$$

Recall the product formula:

$$\sum_{v\in M_K^\infty} \log |a_d|_v^{d_v} + \sum_{v\in M_K^0} \log |a_d|_v^{d_v} = 0.$$

Using this,

$$\begin{aligned}
[K:\mathbb{Q}]\cdot h(\alpha) &= \sum_{v\in M_K^\infty} \sum_{\sigma\in \mathrm{Gal}(K/\mathbb{Q})} \log^+ |\sigma\alpha|_v^{d_v} + \sum_{v\in M_K^\infty} \frac{[K:\mathbb{Q}]}{d} \log |a_d|_v^{d_v} \\
&= \sum_{v\in M_K^\infty} \sum_{1\le j\le d} \frac{[K:\mathbb{Q}]}{d} \log^+ |\alpha_j|_v^{d_v} + \sum_{v\in M_K^\infty} \frac{[K:\mathbb{Q}]}{d} \log |a_d|_v^{d_v}.
\end{aligned}$$

Now $[K:\mathbb{Q}]$ cancels out (which is not surprising since K was just an arbitrary field containing α) and we get

$$\begin{aligned}
d\cdot h(\alpha) &= \sum_{v\in M_K^\infty} \sum_{1\le j\le d} \log^+ |\alpha_j|_v^{d_v} + \sum_{v\in M_K^\infty} \log |a_d|_v^{d_v} \\
&= \sum_{1\le j\le d} \log^+ |\alpha_j| + \log |a_d| = \log M(f_\alpha).
\end{aligned}$$

The last equality is by Jensen's formula (2). □

5 Northcott's property

We are going to show the following.

Theorem 5.1 (Northcott's property). *There are finitely many numbers $\alpha \in \overline{\mathbb{Q}}$ having bounded degree and height.*

Of course bounding both height and degree is essential: if we bound only the height, then we can find an infinite sequence of numbers, e.g.

$$2, \sqrt{2}, \sqrt[3]{2}, \sqrt[4]{2}, \sqrt[5]{2}, \dots$$

The height of these numbers is bounded, since $h(\sqrt[n]{2}) = \frac{1}{n} h(2) \le \log 2$.

Proof of the Northcott's property. Let $\alpha \in \overline{\mathbb{Q}}$ be a number of degree d. Let $h(\alpha) \le \log H$ for some number $H \ge 1$. Consider the minimal polynomial

$$f_\alpha(T) = a_d T^d + \dots + a_1 T + a_0 \in \mathbb{Z}[T].$$

Using the Bézout identities and the triangle inequality we can bound absolute values of the coefficients of f_α:

$$\left|\frac{a_{d-n}}{a_d}\right| = \left|\sum_{1 \le i_1 < \dots < i_n \le d} \alpha_{i_1} \cdots \alpha_{i_n}\right| \le \binom{d}{n} \prod_{1 \le i \le n} \max\{1, |\alpha_i|\}.$$

From the Jensen's formula (2) we obtain

$$|a_{d-n}| \le \binom{d}{n} M(f_\alpha) \le 2^d H^d \quad \text{for all } 0 \le n \le d,$$

where the last inequality comes from $h(\alpha) = \frac{1}{d} \log M(f_\alpha)$. This bound means that one has finitely many choices for the coefficients of f_α, hence finitely many α. □

Namely, from the proof, there are $2(2H)^d + 1$ possibilities for each a_i, and so $(d+1)(2(2H)^d + 1)$ possibilities for f_α, and finally $d(d+1)(2(2H)^d + 1)$ possibilities for picking a root of f_α. This is of course an exhaustive counting, but it is close to the reality (cf. theorem 8.1 below).

Remark 5.2. The proof above is essentially the same what we did for Kronecker's theorem (theorem 2.4). And of course Kronecker's theorem is just a special case: if $h(\alpha) = 0$, then $h(\alpha^n) = n\, h(\alpha) = 0$ for all n, and so the sequence $\alpha, \alpha^2, \alpha^3, \dots$ should consist of finitely many numbers according to the Northcott's property.

6 Lehmer's conjecture and Dobrowolski theorem

Definition 6.1. For a number $\alpha \in \overline{\mathbb{Q}} \setminus \{0\}$ consider its minimal polynomial $f_\alpha \in \mathbb{Z}[T]$. We say that $M(\alpha) := M(f_\alpha)$ is the **Mahler measure** of α.

Remark 6.2. One should be careful: we define $M(\alpha) := M(f_\alpha)$ and we know that $h(\alpha) = \frac{1}{\deg \alpha} \log M(f_\alpha)$. However, the height of the minimal polynomial f_α is not the same as the height of α, as we already observed.

The **Lehmer's conjecture** is the following question: *can we find a constant C such that $M(\alpha) \ge C > 1$ for any $\alpha \in \overline{\mathbb{Q}}$ which is not 0 and not a root of unity?*

For the heights it is equivalent to ask for a constant C' such that

$$h(\alpha) \geq \frac{C'}{d} > 0.$$

This question is easy to pose, since it concerns just heights of *numbers*, not points on abelian varieties or other sophisticated stuff. However, the conjecture is still open in its full generality.

One result in this direction is due to Smyth.

Theorem 6.3 (Smyth, 1971). *Let α be a nonzero algebraic number which is not a root of unity. Assume that its minimal polynomial is not reciprocal (recall that "reciprocal" means $X^d f_\alpha(\frac{1}{X}) = f_\alpha(X)$, or that the coefficients are palindromic). Then $M(\alpha) \geq M(X^3 - X - 1) = 1.324717957\ldots$*

The reciprocal polynomial with the least known Mahler measure is

$$X^{10} - X^9 + X^7 - X^6 + X^5 - X^4 + X^3 - X + 1.$$

It has Mahler measure $1.176280818\ldots$ and it is widely believed to be the least possible value.

The best known bound is due to Dobrowolski.

Theorem 6.4 (Dobrowolski, 1978). *Let α be a nonzero algebraic number which is not a root of unity. Then*

$$M(\alpha) \geq 1 + C\left(\frac{\log\log d}{\log d}\right)^3,$$

where C is a constant not depending on α.

Similarly for the height one has

$$h(\alpha) \geq \frac{C}{d}\left(\frac{\log\log d}{\log d}\right)^3.$$

(Recall that $h(\alpha) = \frac{1}{d}\log M(\alpha)$; and $\log(1+x) \xrightarrow{x\to 0} x$.)

Note that we are interested in the bound for d big enough, since if both $h(\alpha)$ and d are bounded, then there are finitely many such α by the Northcott's property.

Let $f_\alpha(T) = a_d T^d + \cdots + a_1 T + a_0 \in \mathbb{Z}[T]$ be the minimal polynomial of α.

As we observed before, we have a lower bound (3) giving $M(f_\alpha) \geq |a_0|$. If $|a_0| \geq 2$, then we have a nontrivial lower bound $h(\alpha) \geq \frac{1}{d}\log 2$, which is much stronger than the claimed. Similarly, since $M(\alpha) = |a_d| \cdot \prod_{|\alpha_i|\geq 1} |\alpha_i|$, we have $M(\alpha) \geq |a_d|$, and so we are done if $|a_d| \geq 2$. So we may assume $a_d = 1$ and $|a_0| = 1$.

If $\alpha_1, \ldots, \alpha_d$ are the conjugates of α, then we get

$$|\alpha|^d = \left|\prod_{1\leq i\leq d} \alpha_i\right| = |a_0| = 1,$$

so that $|\alpha| = 1$. Hence the only interesting case is when α is an algebraic integer with $|\alpha| = 1$, which is not a root of unity.

☹☹☹ The proof is intricate and can be found in [Bombieri–Gubler, §4.4].

7 Heights on the projective space

Let K be a number field. We have the projective space $\mathbb{P}^N(K)$ with homogeneous coordinates of points $P = (x_0(P) : \cdots : x_N(P))$.

Definition 7.1. For a point $(x_0(P) : \cdots : x_N(P)) \in \mathbb{P}^N(K)$ its **height** is given by

$$h(P) := \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le i \le N} |x_i(P)|_v,$$

where $d := [K : \mathbb{Q}]$ and $d_v := [K_v : \mathbb{Q}_v]$.

We need to check that this is well-defined. If we multiply x_i's by some $\lambda \in K^\times$, then we get by the product formula

$$\begin{aligned}
&\frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le i \le N} |\lambda\, x_i(P)|_v \\
=&\frac{1}{d} \sum_{v \in M_K} d_v \left(\log |\lambda|_v + \log \max_{0 \le i \le N} |x_i(P)|_v \right) \\
=&\frac{1}{d} \log \underbrace{\prod_{v \in M_K} |\lambda|_v^{d_v}}_{=1} + \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le i \le N} |x_i(P)|_v \\
=&\frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le i \le N} |x_i(P)|_v.
\end{aligned}$$

We note that in basic properties like this one the product formula is crucial, and that's why heights of points are defined over global fields.

- We have $h(P) \ge 0$. To see this, we pick a coordinate $x_i(P) \ne 0$ and write the point as $P = \left(\frac{x_0(P)}{x_i(P)} : \cdots : 1 : \cdots : \frac{x_N(P)}{x_i(P)} \right)$. Then

$$h(P) = \frac{1}{d} \sum_{v \in M_K} d_v \log \underbrace{\max_{0 \le i \le N} |x_i(P)|_v}_{\ge 1} \ge 0.$$

- The factor $\frac{1}{d}$ in the definition ensures that the height is well-defined on $\mathbb{P}^N(\overline{\mathbb{Q}})$, i.e. it behaves well under field extensions. For a point $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$ one can take any number field K containing the coordinates of P.
- Observe also that for any $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ one has $h(\sigma(P)) = h(P)$ (cf. proposition 1.6).

Note that the height of a number $\alpha \in \overline{\mathbb{Q}}$ introduced above is the same as the height of the point $(1 : \alpha)$ on the projective line $\mathbb{P}^1(\overline{\mathbb{Q}})$. So on $\mathbb{P}^1$ we have the Northcott property (theorem 5.1) and Kronecker's theorem (theorem 2.4). In fact this is valid also on $\mathbb{P}^N$.

For a point $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$ we denote by $\mathbb{Q}(P)$ a number field containing the coordinates of P.

Theorem 7.2 (Northcott's property). *For any B and D the set*

$$\{P \in \mathbb{P}^N(\overline{\mathbb{Q}}) \mid h(P) \le B \text{ and } [\mathbb{Q}(P) : \mathbb{Q}] \le D\}$$

is finite.

Proof. For a point $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$, up to permuting and normalizing the coordinates, we can assume that $x_0(P) = 1$. Then

$$\begin{aligned}
h(P) :=& \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le i \le N} |x_i(P)|_v \\
\ge& \frac{1}{d} \sum_{v \in M_K} d_v \log \max\{1, |x_i(P)|_v\} \quad \text{(for all } 1 \le i \le N) \\
=& \frac{1}{d} \sum_{v \in M_K} d_v \log^+ |x_i(P)|_v.
\end{aligned}$$

So a bound on $h(P)$ implies a bound on the height $h(x_i(P))$ of each coordinate. Together with a bound on degree $[\mathbb{Q}(P):\mathbb{Q}]$, this allows to apply the Northcott property for heights of numbers (theorem 5.1) and conclude that there are finitely many possibilities for each $x_i(P)$. □

Similarly to the proof above, one deduces using the Kronecker's theorem for numbers (theorem 2.4) its analogue on $\mathbb{P}^N$.

Theorem 7.3 ("Kronecker"). *Let $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$ be a point with $x_i(P) \neq 0$. Then*

$$h(P) = 0 \iff x_j(P) = 0 \text{ or } \frac{x_j(P)}{x_i(P)} \text{ is a root of unity for all } j.$$

8 Schanuel's theorem

We can count the points on the projective space in the following way. For a number field K and a positive parameter $T > 0$ we let

$$\mathcal{N}(\mathbb{P}^n(K), T) := \#\{P \in \mathbb{P}^n(K) \mid H(P)^{[K:\mathbb{Q}]} \leq T\}.$$

This is a finite number, since the condition $H(P)^{[K:\mathbb{Q}]} \leq T$ implies a bound on height (and the degree of K is fixed).

Theorem 8.1 (Schanuel, 1979). *One has asymptotically, for T big enough*

$$\mathcal{N}(\mathbb{P}^n(K), T) = a(K,n)\, T^{n+1} + \begin{cases} O(T \log T), & \text{if } K = \mathbb{Q},\ n = 1, \\ O(T^{n+1-1/d}), & \text{otherwise.} \end{cases}$$

Here $d := [K:\mathbb{Q}]$, and $a(K,n)$ is a constant putting together the arithmetic invariants of K:

$$a(K,n) = \frac{h_K R_K}{w_K \zeta_K(n+1)} \left(\frac{2^{r_1} (2\pi)^{r_2}}{\sqrt{|D_K|}} \right)^{n+1} (n+1)^{r_1+r_2-1},$$

where h_K is the class number, R_K is the regulator, D_K is the discriminant, w_K is the number of roots of unity, ζ_K is the Dedekind zeta function defined by

$$\zeta_K(s) = \sum_{I \subseteq \mathcal{O}_K} \frac{1}{N(I)^s} \quad (\mathrm{Re}(s) > 1),$$

r_1 is the number of real places and r_2 is the number of conjugate pairs of complex places (so that $d = r_1 + 2r_2$).

This theorem shows that counting points of bounded height is closely related to the arithmetic. So heights give a link between geometry and arithmetic. This is the main idea of the course.

Proof idea. We focus on the easiest case $K = \mathbb{Q}$. One has $h_{\mathbb{Q}} = R_{\mathbb{Q}} = D_{\mathbb{Q}} = 1$, $w_{\mathbb{Q}} = 2$, $r_1 = 1$, $r_2 = 0$, and the constant is

$$a(\mathbb{Q}, n) = \frac{2^n}{\zeta_K(n+1)}.$$

For a point $P = (x_0 : \cdots : x_n) \in \mathbb{P}^n(\mathbb{Q})$ we normalize the coordinates in a way that $x_0, \ldots, x_n \in \mathbb{Z}$ and $\gcd(x_0, \ldots, x_n) = 1$. This representation of a point is unique up to sign $\pm$. So when we count such points we will divide the total number by two.

In this case the height $H(P)$ is just the maximal value $|x_i|$ for $i = 0, \dots, n$. Instead of counting points on $\mathbb{P}^n$, we will count points on $\mathbb{A}^{n+1}$ with the described normalization. For a point $\underline{x} = (x_0, \dots, x_n) \in \mathbb{A}^{n+1}(\mathbb{Z})$ we put

$$|\underline{x}| := \max_{0 \le i \le n} |x_i|,$$
$$\gcd \underline{x} := \gcd(x_0, \dots, x_n).$$

Let us define

$$M(T) := \#\{\underline{x} \in \mathbb{A}^{n+1}(\mathbb{Z}) \mid \underline{x} \neq \underline{0},\ |\underline{x}| \le T\},$$
$$M^*(T, \delta) := \#\{\underline{x} \in \mathbb{A}^{n+1}(\mathbb{Z}) \mid \gcd \underline{x} = \delta,\ |\underline{x}| \le T\}.$$

Thus we are interested in

$$\mathcal{N}(\mathbb{P}^n(\mathbb{Q}), T) = \frac{1}{2} M^*(T, 1),$$

where the factor $\frac{1}{2}$ comes from the sign choice, as we said above.

Observe that $M^*(T, \delta) = M^*(\frac{T}{\delta}, 1)$, hence

$$M(T) = \sum_{\delta \ge 1} M^*(T, \delta) = \sum_{\delta \ge 1} M^*\left(\frac{T}{\delta}, 1\right).$$

The latter is a finite sum, since $M^*(\frac{T}{\delta}, 1) = 0$ if $\delta > T$.

We will make use of the Möbius inversion formula.

Remark 8.2. Recall that the **Möbius function** is given by

$$\mu(p_1 \cdots p_r) := (-1)^r, \quad p_i \text{ are distinct primes},$$
$$\mu(1) := 1,$$
$$\mu(x) := 0 \quad \text{otherwise (if there are squared factors)}.$$

The fundamental property of μ is that for $n > 1$

$$\sum_{d \mid n} \mu(d) = 0.$$

Recall that from this follows the **Möbius inversion formula**: if F and F^* are functions satisfying

$$F(T) = \sum_{1 \le \delta \le T} F^*\left(\frac{T}{\delta}\right) \quad \text{for all } T \ge 1,$$

then one has

$$F^*(T) = \sum_{1 \le \delta \le T} \mu(\delta) F\left(\frac{T}{\delta}\right) \quad \text{for all } T \ge 1.$$

In our case the Möbius inversion formula gives

$$M^*(T, 1) = \sum_{\delta \ge 1} \mu(\delta) M\left(\frac{T}{\delta}\right).$$

We have

$$M\left(\frac{T}{\delta}\right) = \#\{\underline{x} \in \mathbb{A}^{n+1}(\mathbb{Z}) \mid \underline{x} \neq \underline{0},\ |\underline{x}| \le \frac{T}{\delta}\} = \left(2\left[\frac{T}{\delta}\right] + 1\right)^{n+1} - 1.$$

For T big enough we can write

$$\begin{aligned} M^*(T,1) &= \sum_{1\le\delta\le T} \mu(\delta)\left(\frac{2T}{\delta}+O(1)\right)^{n+1} \\ &= \sum_{1\le\delta\le T} \mu(\delta)\left(\left(\frac{2T}{\delta}\right)^{n+1}+O\left(\frac{T}{\delta}\right)^{n+1}\right) \\ &= (2T)^{n+1}\sum_{\delta\ge 1}\mu(\delta)\,\delta^{-(n+1)} - (2T)^{n+1}\sum_{\delta>T}\mu(\delta)\,\delta^{-(n+1)} + O\left(T^n \sum_{1\le\delta\le T}\delta^{-n}\right). \end{aligned}$$

Now for the zeta function $\zeta(s)=\sum_{n\ge1}\frac{1}{n^s}$ the series for $\zeta(s)^{-1}$ is given by $\sum_{n\ge1}\frac{\mu(n)}{n^s}$. Using this, we can write the latter sum as

$$(2T)^{n+1}\frac{1}{\zeta(n+1)} + \begin{cases} O(T\log T), & \text{if } n=1,\\ O(T^{n+1}), & \text{otherwise.}\end{cases}$$

To obtain $\mathcal{N}(\mathbb{P}^n(\mathbb{Q}),T)$ we should divide this by two, and we see that this is what we wanted to show. □

For a projective variety $V\subset\mathbb{P}^n$ (with some fixed embedding in $\mathbb{P}^n$; see more on this below) we can similarly count the number of points

$$\mathcal{N}(V(K),T) := \#\{P\in V(K)\mid H(P)^{[K:\mathbb{Q}]}\le T\}.$$

It is *very* difficult to estimate $\mathcal{N}(V(K),T)$, but for curves ($\dim V=1$) the answer is known. Namely, if C is a smooth curve defined over a number field K, then, assuming $C(K)\neq\varnothing$, one has

- If C is a rational curve (of genus 0), then for some $a>0$, $b>0$
$$\mathcal{N}(C(K),T)\sim a\,T^b.$$
- If C is an elliptic curve (of genus 1), then for some $a>0$, $b\ge0$
$$\mathcal{N}(C(K),T)\sim a\,(\log T)^b.$$
- If C is of genus ≥2, then $C(K)$ is finite.

The last case is known as the **Mordell conjecture**, which was solved by Gerd Faltings in 1983. We will come back to this later on.

9 Divisors

Here we briefly summarize some facts about divisors, since they will be very important starting from the next section. (There was no lecture on this, but this material is useful for the reference.)

For simplicity, we let V/K be a smooth projective algebraic variety.

Definition 9.1. The group of **Weil divisors** $\mathrm{Div}(V)$ is the free abelian group generated by the closed irreducible subvarieties of codimension one in V.

For a Weil divisor $D=\sum n_Y Y$ with $n_Y\in\mathbb{Z}$ its **support** $\operatorname{supp} D$ is the subvariety $\bigcup_{n_Y\neq0}Y$.

D is called an **effective divisor** if $n_Y\ge0$.

The **degree** of $D=\sum n_Y Y$ is the number $\sum n_Y\in\mathbb{Z}$.

If $\eta\in Y$ is the generic point, then the local ring $\mathcal{O}_{V,\eta}$ is a discrete valuation ring with quotient field $K(V)$, and so it defines valuation $v_Y\colon K(V)^\times\to\mathbb{Z}$.

Definition 9.2. For a function $f \in K(V)^\times$ we define

$$(f) := \sum_Y v_Y(f)\, Y = \boxed{\text{zeros with multiplicities}} - \boxed{\text{poles with multiplicities}}$$

(the sum is well-defined, since $v_Y(f) \neq 0$ only for finitely many Y).

Such a Weil divisor (f) is called **principal**.

Two Weil divisors D_1, D_2 are called **linearly equivalent** ($D_1 \sim D_2$) if $D_1 - D_2$ is a principal divisor.

Definition 9.3. A **Cartier divisor** on a variety V is defined by a an open cover $V = \bigcup_{i\in I} U_i$ and functions $f_i \in K(U_i)^\times = K(V)^\times$ that satisfy the compatibility condition

$$f_i f_j^{-1} \in \mathcal{O}(U_i \cap U_j)^\times \quad \text{for all } i, j \in I,$$

i.e. $f_i f_j^{-1}$ has no poles or zeros on $U_i \cap U_j$.

Further, two such collections $\{(U_i, f_i)\}$ and $\{(W_j, g_j)\}$ are equivalent (define the same divisor) is for all $i \in I, j \in J$ one has $f_i g_j^{-1} \in \mathcal{O}(U_i \cap W_j)^\times$.

The sum of two Cartier divisors is given by

$$\{(U_i, f_i)\} + \{(W_j, g_j)\} := \{(U_i \cap W_j, f_i g_j)\}.$$

With this operation Cartier divisors on V form a group $\mathrm{DivC}(V)$.

A Cartier divisor is called **effective** if it can be defined by $\{(U_i, f_i)\}$ with $f_i \in \mathcal{O}(U_i)$.

Definition 9.4. A Cartier divisor $\mathcal{D}$ is called **principal** if for some $f \in K(V)^\times$ one has $\mathcal{D} = \{(V, f)\}$.

Two Cartier divisors $\mathcal{D}_1, \mathcal{D}_2$ are called **linearly equivalent** ($\mathcal{D}_1 \sim \mathcal{D}_2$) if $\mathcal{D}_1 - \mathcal{D}_2$ is a principal divisor.

Let $\{(U_i, f_i)\}$ define a Cartier divisor on V. For an irreducible subvariety of codimension one Y we pick U_i such that $U_i \cap Y \neq \varnothing$, and we consider $v_Y(f_i)$. This does not depend on the choice of U_i and gives a homomorphism

$$\begin{aligned} &\mathrm{DivC}(V) \to \mathrm{Div}(V), \\ &\{(U_i, f_i)\} \mapsto \sum_Y v_Y(f_i) \end{aligned}$$

Fact 9.5 (Hartshorne II.6.11). *If V is a smooth variety, then the map $\mathrm{DivC}(V) \to \mathrm{Div}(V)$ is an isomorphism, and it induces an isomorphism of groups*

$$\mathrm{Pic}(V) := \frac{\textit{Cartier divisors on } V}{\textit{linear equivalence}} \simeq \frac{\textit{Weil divisors on } V}{\textit{linear equivalence}}.$$

Definition 9.6. The group $\mathrm{Pic}(V)$ above is called the **Picard group** of V.

Example 9.7. For $V = \mathbb{P}^n$ let $H \subset \mathbb{P}^n$ be the hyperplane $x_0 = 0$. Then for any divisor $D \in \mathrm{Div}(\mathbb{P}^n)$ of degree d one has $D \sim d\,H$. The degree gives an isomorphism $\mathrm{Pic}(\mathbb{P}^n) \to \mathbb{Z}$.

We will work interchangeably with Weil and Cartier divisors, since our varieties are always smooth.

Definition 9.8. For a morphism of varieties $g\colon V \to W$ let $\mathcal{D} = \{(U_i, f_i)\}$ define a Cartier divisor on W. Up to replacing the divisor within its linear equivalence class ("moving lemma"), one can assume that $g(V) \not\subset \operatorname{supp}\mathcal{D}$. The Cartier divisor $g^*(\mathcal{D})$ is the Cartier divisor on V given by

$$g^*(\mathcal{D}) := \{(g^{-1}(U_i), f_i \circ g)\}.$$

This induces a homomorphism $g^*\colon \mathrm{Pic}(W) \to \mathrm{Pic}(V)$.

One sees that Pic is a contravariant functor.

Definition 9.9. For a divisor $D \in \mathrm{Div}(V)$ we define the K-vector space

$$L(D) := \{f \in K(V)^\times \mid D + (f) \geq 0\} \cup \{0\}$$

(this is indeed a vector space since $v_Y(f+g) \geq \min\{v_Y(f), v_Y(g)\}$).

We denote $\ell(D) := \dim_K L(D)$.

If V is a projective variety (which we always assume), then $L(D)$ is finite dimensional.

Definition 9.10. A **linear system** on a variety V is a set of effective divisors all linearly equivalent to a fixed divisor D, parametrized by a linear subvariety of $\mathbb{P}(L(D)) \simeq \mathbb{P}^{\ell(D)-1}$.

Definition 9.11. For a divisor $D \in \mathrm{Div}(V)$ the set $|D|$ of all effective divisors that are linearly equivalent to D is parametrized by

$$\begin{aligned} \mathbb{P}(D(L)) &\to |D|, \\ f \pmod{K^\times} &\mapsto D + (f). \end{aligned}$$

We call $|D|$ the **complete linear system** of D.

Definition 9.12. Let L be a linear system of dimension n parametrized by a projective space $\mathbb{P}(V) \subset \mathbb{P}(L(D))$. Fix a basis $f_0, \ldots, f_n$ of $V \subset L(D)$. The **rational map associated to** L is given by

$$\begin{aligned} \phi_L \colon V &\dashrightarrow \mathbb{P}^n, \\ x &\mapsto (f_0(x) : \cdots : f_n(x)). \end{aligned}$$

(This "definition" is very wrong and depends on the basis choice, but eventually for our applications it won't be a problem.)

The set of **base points** of L is the intersection of supports of all divisors in L. The rational map ϕ_L is defined outside of the base points of L.

A divisor is called **base point free** if $|D|$ has no base points.

Definition 9.13. A linear system L on a projective variety V is called **very ample** if the rational map $\phi_L \colon V \dashrightarrow \mathbb{P}^n$ is an embedding (= it is actually a morphism, and it maps V isomorphically onto its image $\phi_L(V)$).

A divisor D is called **very ample** if $|D|$ is very ample.

A divisor D is called **ample** if $\underbrace{D + \cdots + D}_{n}$ is very ample for some n.

Fact 9.14 (Hartshorne, Exercise III.5.7)**.** *Let $f \colon V \to W$ be a morphism of projective varieties. If $D \in \mathrm{Div}(W)$ is base point free (resp. ample), then $f^*D \in \mathrm{Div}(V)$ is base point free (resp. ample).*

Fact 9.15. *Let D be any divisor and let H be an ample divisor. Then $D + mH$ is base point free for m big enough. If D is base point free, then $D + H$ is very ample.*

Corollary 9.16. *Every divisor D can be written as a difference of two base point free (very) ample divisors.*

Proof. Pick H a very ample divisor. For m big enough, $D + mH$ and mH are base point free very ample, and we write

$$D = (D + mH) - mH.$$

□

10 Heights on projective varieties

We would like to study heights not just on the points of $\mathbb{P}^N$, but on the points of projective varieties. The first interesting projective variety, apart from $\mathbb{P}^N$ itself, is $\mathbb{P}^n \times \mathbb{P}^m$. Recall that we have the **Segre embedding**

$$\begin{aligned} S_{n,m}\colon \mathbb{P}^n(K) \times \mathbb{P}^m(K) &\to \mathbb{P}^N(K), \\ (x, y) &\mapsto (x_0 y_0 : x_0 y_1 : \cdots : x_i y_j : \cdots : x_n y_m). \end{aligned}$$

Here $N = (n+1)(m+1) - 1$.

Proposition 10.1. *Let K be a number field. For any $x \in \mathbb{P}^n(K)$ and $y \in \mathbb{P}^m(K)$ one has*

$$h(S_{n,m}(x, y)) = h(x) + h(y).$$

Proof. This is immediate:

$$\begin{aligned} h(S_{n,m}(x, y)) &= \sum_{v \in M_K} \frac{d_v}{d} \log \max_{\substack{1 \le i \le n \\ 1 \le j \le m}} |x_i y_j|_v \\ &= \sum_{v \in M_K} \frac{d_v}{d} \log \max_{1 \le i \le n} |x_i|_v + \sum_{v \in M_K} \frac{d_v}{d} \log \max_{1 \le j \le m} |y_j|_v \\ &= h(x) + h(y). \end{aligned}$$

□

If we have a projective variety V/K that embeds in $\mathbb{P}^n$, then we can consider heights of the points $V(K)$ viewed as points in $\mathbb{P}^n(K)$.

Definition 10.2. For a morphism $\phi\colon V \to \mathbb{P}^n$ the **height with respect to** ϕ is given by

$$h_\phi(P) := h(\phi(P)) \quad \text{for } P \in V(\overline{\mathbb{Q}}).$$

However, a priori this is not very interesting, since there are various embeddings, so we should make sure the height does not depend seriously on that. The following preliminary result tells how a rational map $\mathbb{P}^n \dashrightarrow \mathbb{P}^m$ changes the height.

Proposition 10.3. *Let $\phi = (f_0 : \cdots : f_m)\colon \mathbb{P}^n \dashrightarrow \mathbb{P}^m$ be a rational map given by $m+1$ homogeneous polynomials $f_i(X_0 : \cdots : X_n)$ of degree d. Let $Z \subset \mathbb{P}^n$ be the set of common zeros of $f_0, \ldots, f_m$. Then*

1. *For any point $P \in \mathbb{P}^n(\overline{\mathbb{Q}}) \setminus Z$ one has*

$$h(\phi(P)) \le d\, h(P) + C,$$

   where C is a constant not depending on P (but depending on ϕ).

2. *If X is a closed subvariety in $\mathbb{P}^n$ and $X \cap Z = \varnothing$, then*

$$h(\phi(P)) \ge d\, h(P) + C' \quad \text{for } P \in X(\overline{\mathbb{Q}})$$

   for some constant C' not depending on P.

Proof. Each homogeneous polynomial $f_i(X_0 : \cdots : X_n)$ can be written as

$$f_i(X_0 : \cdots : X_n) = \sum_{e_0 + \cdots + e_n = d} a_{i, e_0, \ldots, e_n} X_0^{e_0} \cdots X_n^{e_n}.$$

Totally there are $\binom{n+d}{n}$ monomials in the sum.

Now we have by definition

$$h(\phi(P)) = \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le i \le m} |f_i(P)|_v.$$

We want to estimate

$$|f_i(P)|_v = \left| \sum_{e_0 + \cdots + e_n = d} a_{i,e_0,\ldots,e_n} x_0(P)^{e_0} \cdots x_n(P)^{e_n} \right|_v .$$

If v is nonarchimedian, then we just take

$$|f_i(P)|_v \le \max_{e_0,\ldots,e_n} |a_{i,e_0,\ldots,e_n}|_v \cdot \max_{0 \le j \le n} |x_j(P)|_v^d.$$

If v is archimedian, then we must use the triangle inequality. Since there are $\binom{n+d}{n}$ monomials, the right hand side will be the same, but with an extra factor $\binom{n+d}{n}$.

Now we take logarithms and sum over all $v \in M_K$. Observe that we can drop the part with $a_{i,e_0,\ldots,e_n}$, and also the part $\binom{n+d}{n}$ appearing for the archimedian places, since they sum up to some constant depending only on ϕ. What remains is

$$h(\phi(P)) \le \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le j \le n} |x_j(P)|_v^d + C.$$

We recognize $d\,h(P)$, and the rest is some constant not depending on P, so we proved the first part.

For the second part we need to use some geometry. Assume that the closed subvariety X in $\mathbb{P}^n$ is defined by some homogeneous polynomials $(p_1, \ldots, p_r) = I(X)$. By assumption they have no common zeros with $f_0, \ldots, f_m$, hence by the Nullstellensatz

$$\sqrt{(p_1, \ldots, p_r, f_0, \ldots, f_m)} = I(\varnothing) = (X_0, \ldots, X_n).$$

That is, there exists an exponent $t \ge d$ such that each X_j can be written as

$$X_j^t = q_{j,1} \cdot p_1 + \cdots + q_{j,r} \cdot p_r + g_{j,0} \cdot f_0 + \cdots + g_{j,m} \cdot f_m.$$

Here $q_{j,k}$ and $g_{j,\ell}$ are some polynomials. A priori this works only over algebraic closure, but by passing to a finite extension we may assume the coefficients of these polynomials lie in K. Now if we take a point $P \in X(K)$, then $p_k(P) = 0$, and

$$x_j(P)^t = g_{j,0}(P) \cdot f_0(P) + \cdots + g_{j,m}(P) \cdot f_m(P).$$

Here $g_{j,i}$ are some homogeneous polynomials of degree $t - d$. We can use this to estimate

$$\begin{aligned}
t\,h(P) &= \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le j \le n} |x_j(P)|_v^t \\
&= \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le j \le n} |g_{j,0}(P) \cdot f_0(P) + \cdots + g_{j,m}(P) \cdot f_m(P)|_v \\
&\le \frac{1}{d} \sum_{v \in M_K^0} d_v \log \max_{\substack{0 \le j \le n \\ 0 \le i \le m}} |g_{j,i}(P) \cdot f_i(P)|_v + \frac{1}{d} \sum_{v \in M_K^\infty} d_v \log((m+1) \max_{\substack{0 \le j \le n \\ 0 \le i \le m}} |g_{j,i}(P) \cdot f_i(P)|_v) \\
&\le \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{\substack{0 \le j \le n \\ 0 \le i \le m}} |g_{j,i}(P)|_v + \frac{1}{d} \sum_{v \in M_K} d_v \log \max_{0 \le i \le m} |f_i(P)|_v + O(1)
\end{aligned}$$

Since $g_{j,i}$ are homogeneous polynomials of degree $t - d$, we bound the first term by $\le (t-d)\,h(P) + O(1)$; the second term is exactly $h(\phi(P))$, and remains the desired inequality

$$d\,h(P) \le h(\phi(P)) + O(1). \qquad \square$$

Given $\phi: V \to \mathbb{P}^n$, we can study the height $h_\phi(P) := h(\phi(P))$. However, it is important to know how h_ϕ is related to h_ψ if we take another morphism $\psi: V \to \mathbb{P}^m$.

Proposition 10.4. *Let V be a projective variety defined over $\overline{\mathbb{Q}}$. Let $H \subset \mathbb{P}^n$ and $H' \subset \mathbb{P}^m$ be hyperplanes. Let $\phi: V \to \mathbb{P}^n$ and $\psi: V \to \mathbb{P}^m$ are morphisms such that $\phi^* H$ and $\psi^* H'$ are linearly equivalent (meaning that the morphisms ϕ and ψ come from the same complete linear system). Then one has*

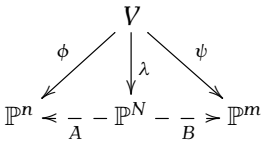
$$h_\phi(P) = h_\psi(P) + O(1),$$

where $O(1)$ depends on everything but not on $P \in V(\overline{\mathbb{Q}})$.

Proof. Let D be an effective divisor in the linear equivalence class of $\phi^* H$ and $\psi^* H'$. If $h_0, \dots, h_N$ is a basis of the vector space $L(D)$, then we have $\phi = (f_0 : \cdots : f_n)$ and $\psi = (g_0 : \cdots : g_m)$, and

$$f_i = \sum_{0 \le j \le N} a_{ij} h_j,$$
$$g_i = \sum_{0 \le j \le N} b_{ij} h_j.$$

Let $\lambda = (h_0 : \cdots : h_N): V \to \mathbb{P}^N$ be the morphism corresponding to the complete linear system $|D|$. Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices giving rational maps $A: \mathbb{P}^N \dashrightarrow \mathbb{P}^n$ and $B: \mathbb{P}^N \dashrightarrow \mathbb{P}^m$. We get a commutative diagram



Here A and B are defined not on all $\mathbb{P}^N$, but they are defined on the image of λ.
Now using proposition 10.3 (in the special case $d = 1$), we have

$$h(\phi(P)) = h(A(\lambda(P))) = h(\lambda(P)) + O(1),$$
$$h(\psi(P)) = h(B(\lambda(P))) = h(\lambda(P)) + O(1),$$

so we are done. □

11 Weil's height machine

For any smooth projective variety V defined over a number field K, we can take a divisor D on V and produce from it a height function $h_{V,D}: V(\overline{K}) \to \mathbb{R}$. We state as a theorem all the desired properties of $h_{V,D}$.

Theorem 11.1. *There exists a map*

$$h_V: \mathrm{Div}(V) \to \{\text{functions } V(\overline{K}) \to \mathbb{R}\}$$

which is up to a constant $O(1)$ uniquely defined by the following three properties:

- ***Normalization**: for a hyperplane of codimension one $H \subset \mathbb{P}^n$ one has for all $P \in \mathbb{P}^n(\overline{K})$*

  $$h_{\mathbb{P}^n, H}(P) = h(P) + O(1),$$

  where on the right hand side is the usual height on $\mathbb{P}^n(\overline{K})$.

- ***Functoriality**: for a morphism of varieties $\phi: V \to W$ inducing a map $\phi^*: \mathrm{Div}(W) \to \mathrm{Div}(V)$, and a divisor $D \in \mathrm{Div}(W)$ one has for all $P \in V(\overline{K})$*

$$h_{V,\phi^* D}(P) = h_{W,D}(\phi(P)) + O(1).$$

- ***Additivity**: for $D, E \in \mathrm{Div}(V)$ one has for all $P \in V(\overline{K})$*

$$h_{V,D+E}(P) = h_{V,D}(P) + h_{V,E}(P) + O(1).$$

Further, such h_V satisfies the following additional properties:

- ***Linear equivalence**: if $D, E \in \mathrm{Div}(V)$ are two linearly equivalent divisors, then*

$$h_{V,D} = h_{V,E} + O(1).$$

- ***Positivity**: If D is an effective divisor on V, let B be the set of base points of the associated linear system $|D|$. Then for all $P \in (V \setminus B)(\overline{K})$*

$$h_{V,D}(P) \geq O(1).$$

- ***Northcott's property**: if $D \in \mathrm{Div}(V)$ is an ample divisor, then for any finite field extension K'/K and any constant M the set*

$$\{P \in V(K') \mid h_{V,D}(P) \leq M\}$$

is finite.

By $O(1)$ we always mean constants depending on all the datum (the variety, divisors, morphisms) but not on a particular point P.

The construction for D base point free. A divisor D defines a rational map $\phi_{|D|}: V \dashrightarrow \mathbb{P}^n$ (cf. definition 9.12), which is defined outside of the base points of $|D|$. Hence if D is base point free, then we can define for $P \in V(\overline{K})$

$$h_{V,D}(P) := h(\phi_{|D|}(P)).$$

A morphism $\phi_{|D|}: V \to \mathbb{P}^n$ is not uniquely defined, however we saw in proposition 10.4 that up to $O(1)$, the definition above depends only on D.

A hyperplane $H \subset \mathbb{P}^n$ gives rise to the identity morphism $\mathbb{P}^n \to \mathbb{P}^n$, and so the normalization property $h_{\mathbb{P}^n,H}(P) = h(P) + O(1)$ is clear for the definition above.

Additivity for D base point free. For two base point free divisors D and E we consider associated morphisms $\phi_{|D|}: V \to \mathbb{P}^n$ and $\phi_{|E|}: V \to \mathbb{P}^m$. We use the Segre embedding:

$$V \xrightarrow{\phi_{|D|} \times \phi_{|E|}} \mathbb{P}^n \times \mathbb{P}^m \xrightarrow{S_{n,m}} \mathbb{P}^N, \qquad V \xrightarrow{=: \phi_{|D|} \otimes \phi_{|E|}} \mathbb{P}^N$$

Let H_n, H_m, H_N be hyperplanes in $\mathbb{P}^n, \mathbb{P}^m, \mathbb{P}^N$ respectively. Then we have a linear equivalence in $\mathrm{Div}(\mathbb{P}^n \times \mathbb{P}^m)$

$$S_{n,m}^* H_N \sim H_n \times \mathbb{P}^m + \mathbb{P}^n \times H_m.$$

Indeed, if H_N is given by $\{(z_0 : \cdots : z_N) \mid z_0 = 0\}$, and similarly $H_n = \{(x_0 : \cdots : x_n) \mid x_0 = 0\}$, $H_m = \{(y_0 : \cdots : y_m) \mid y_0 = 0\}$, then

$$S_{n,m}^* H_N = \{(x, y) \in \mathbb{P}^n \times \mathbb{P}^m \mid x_0 y_0 = 0\} = H_n \times \mathbb{P}^m + \mathbb{P}^n \times H_m.$$

We have now

$$(\phi_{|D|} \otimes \phi_{|E|})^* H \sim D + E.$$

And so

$$h_{V,D+E}(P) = h((\phi_{|D|} \otimes \phi_{|E|})(P)) + O(1) = h(S_{n,m}(\phi_{|D|}(P), \phi_{|E|}(P))) + O(1).$$

By proposition 10.1, the latter is equal to $h(\phi_{|D|}(P)) + h(\phi_{|E|}(P)) + O(1)$, i.e. to $h_{V,D}(P) + h_{V,E}(P) + O(1)$. Thus we have the additivity property for base point free divisors.

The construction for D an arbitrary divisor. Observe that a very ample divisor D gives rise to an embedding $\phi_{|D|}: V \hookrightarrow \mathbb{P}^n$. By the normalization and functoriality property, we are *forced* to define for very ample divisors

$$h_{V,D} := h \circ \phi_{|D|} + O(1).$$

Further, any divisor D can be written as $D_1 - D_2$ for D_1, D_2 base point free and very ample (corollary 9.16), and by the additivity property we are forced to put

$$h_{V,D}(P) := h_{V,D_1}(P) - h_{V,D_2}(P).$$

We just need to check that different decompositions $D_1 - D_2$ lead to the same height up to constant. If $D_1 - D_2 = E_1 - E_2$, then using the additivity property for the base point free case,

$$h_{V,D_1} + h_{V,E_2} = h_{V,D_1+E_2} + O(1) = h_{V,D_2+E_1} + O(1) = h_{V,D_2} + h_{V,E_1} + O(1),$$

and so $h_{V,D_1} - h_{V,D_2} = h_{V,E_1} - h_{V,E_2} + O(1)$.

Functoriality. For a morphism $\phi: V \to W$ one has

$$h_{V,\phi^* D} = h_{V,\phi^* D_1} - h_{V,\phi^* D_2} + O(1) = h_{W,D_1} \circ \phi - h_{W,D_2} \circ \phi + O(1) = h_{W,D} \circ \phi + O(1).$$

Here we use that $\phi^* D_1$ and $\phi^* D_2$ are base point free as well (fact 9.14).

Additivity. If $D = D_1 - D_2$ and $E = E_1 - E_2$ are corresponding decompositions, then

$$h_{V,D+E} = h_{V,D_1+E_1} - h_{V,D_2+E_2} + O(1) = h_{V,D_1} + h_{V,E_1} - h_{V,D_2} - h_{V,E_2} + O(1) = h_{V,D} + h_{V,E} + O(1).$$

So we defined $h_{V,D}$ for any divisor $D \in \mathrm{Div}(V)$ and we saw that the definition is, up to constants, the only possible if we require normalization, functoriality, and additivity.

Linear equivalence. If $D_1 - D_2 \sim E_1 - E_2$, then by proposition 10.4

$$h(\phi_{|D_1+E_2|}(P)) = h(\phi_{|D_2+E_1|}(P)) + O(1).$$

Hence

$$h_{V,D_1} + h_{V,E_2} = h_{V,D_2} + h_{V,E_1} + O(1),$$

and now by the additivity

$$h_{V,D_1-D_2} = h_{V,D_1} - h_{V,D_2} + O(1) = h_{V,E_1} - h_{V,E_2} + O(1) = h_{V,E_1-E_2} + O(1).$$

Positivity. If D is an effective divisor, we write it as $D_1 - D_2$ where D_1 and D_2 are base point free. Let $f_0, \dots, f_n$ be a basis of $L(D_2)$. Since D is effective,

$$D_1 + (f_i) = D + D_2 + (f_i) \geq 0,$$

hence $f_0, \dots, f_n \in L(D_1)$, meaning that this extends to a basis $f_0, \dots, f_n, f_{n+1}, \dots, f_m$ of $L(D_1)$. With respect to this basis, we have isomorphisms $\phi_{|D_1|}\colon V \to \mathbb{P}^n$ and $\phi_{|D_2|}\colon V \to \mathbb{P}^n$. Now if $P \notin \operatorname{supp}(D_1)$, we have

$$\begin{aligned} h_{V,D}(P) &= h_{V,D_1}(P) - h_{V,D_2}(P) + O(1) \\ &= h(\phi_{|D_1|}(P)) - h(\phi_{|D_2|}(P)) + O(1) \\ &= h(f_0(P) : \dots : f_m(P)) - h(f_0(P) : \dots : f_n(P)) + O(1). \end{aligned}$$

But $m \geq n$, hence $h(f_0(P) : \dots : f_m(P)) \geq h(f_0(P) : \dots : f_n(P))$, which means $h_{V,D}(P) \geq O(1)$.

This shows the positivity $h_{V,D}(P) \geq O(1)$ for the points $P \notin \operatorname{supp} D_1$. In general, we can take H to be a very ample divisor such that $D + H$ is very ample (cf. fact 9.15). This gives an embedding $\phi\colon V \hookrightarrow \mathbb{P}^N$. Let $H_0, \dots, H_N$ be the pullbacks of the coordinate hyperplanes in $\mathbb{P}^N$. Then $H_0 \cap \dots \cap H_N = \varnothing$, each H_i and $H_i + D$ are very ample. We can write $D = (D + H_i) - H_i$ and apply the above to deduce positivity $h_{V,D}(P) \geq O(1)$ for the points $P \notin \operatorname{supp} D$. By varying D in its linear system $|D|$, we deduce $h_{V,D}(P) \geq O(1)$ for all P outside of the set of base points of $|D|$.

Northcott's property. Observe that if D is ample, then $m\,D$ is very ample for some m, and so gives an embedding $\phi_{|mD|}\colon V \hookrightarrow \mathbb{P}^n$, i.e. $\phi^*_{|mD|} H = m\,D$. So by normalization, functoriality, and additivity,

$$m\,h_{v,D} = h_{V,m\,D} + O(1) = h_{V,\phi^*_{|mD|}H} + O(1) = h_{\mathbb{P}^n,H} \circ \phi_{|mD|} + O(1) = h \circ \phi_{|m\,D|} + O(1).$$

The latter is bounded by the Northcott's property on $\mathbb{P}^n$ (theorem 7.2).

12 Néron–Tate height on abelian varieties

Recall that an **abelian variety** is a projective variety that is also an algebraic group. The group law is automatically abelian (hence the name) and the variety and automatically smooth.

Example 12.1. Over $\mathbb{C}$ any abelian variety $(A(\mathbb{C}), H)$ is isomorphic to $(\mathbb{C}^g/\Lambda, H)$, where $g = \dim A$ and Λ is a lattice in $\mathbb{C}^g$ (which can be normalized so that $\Lambda = \mathbb{Z}^g + \tau\,\mathbb{Z}^g$ for a $g \times g$ matrix τ). Further, $H\colon \mathbb{C}^g \times \mathbb{C}^g \to \mathbb{C}^g$ is a positive definite bilinear form satisfying

1. $H(z, w) = \overline{H(w, z)}$,
2. $\operatorname{Im} H$ takes values in $\mathbb{Z}$ on $\Lambda \times \Lambda$.

Example 12.2. For $g = 1$ we have **elliptic curves**. In particular, over $\mathbb{C}$ a form H as above exists for any lattice $\Lambda \subset \mathbb{C}$. Further $\mathbb{C}/\Lambda$ embeds in $\mathbb{P}^2(\mathbb{C})$ via the map $z \mapsto (\wp(z) : \wp'(z) : 1)$. We recall that any elliptic curve can be given in $\mathbb{P}^2(K)$ by the **Weierstrass equation**

$$Y^2 Z = X^3 + a\,X Z^2 + b\,Z^3.$$

Example 12.3. If E is a smooth projective curve of genus one, then picking a point $O \in E(K)$ gives us the group law of elliptic curve. If C is a curve of genus ≥ 2, then C itself has no group law, however it embeds in its **Jacobian**: $C \hookrightarrow \operatorname{Jac}(C)$, which is an abelian variety.

Our goal is to define a height on the points of an abelian variety $A(\overline{K})$ which would be compatible with the group law on $A(\overline{K})$. First we examine what the Weil's height machine gives us.

Proposition 12.4. *Let A/K be an abelian variety over a number field. Let $D \in \operatorname{Div}(A)$ and for $m \in \mathbb{Z}$ denote by $[m]$ the map $P \mapsto \underbrace{P + \dots + P}_{m}$ (with $[m]\,P := -[-m]\,P$ for $m < 0$). Then*

$$h_{A,D}([m]\,P) = \frac{m^2 + m}{2}\,h_{A,D}(P) + \frac{m^2 - m}{2}\,h_{A,D}(-P) + O(1),$$

where the constant does not depend on P.

Proof. This is a particular consequence of the **cube theorem**. If $f, g, h: V \to A$ are regular maps from a variety V to an abelian variety A, then the divisor

$$(f+g+h)^*D - (f+g)^*D - (f+h)^*D - (g+h)^*D + f^*D + g^*D + h^*D$$

is equivalent to O.

If we take $f = [m]$, $g = id$, $h = [-1]$, then we get

$$[m]^*D - [m+1]^*D - [m-1]^*D + [m]^*D + D + [-1]^*D \sim O.$$

In other words,

$$[m+1]^*D \sim 2[m]^*D - [m-1]^*D + D + [-1]^*D. \tag{5}$$

By induction on m, this implies Mumford's formula

$$[m]^*D \sim \frac{m^2+m}{2} D + \frac{m^2-m}{2} [-1]^*D. \tag{6}$$

This is what we need, and the rest follows by additivity, functoriality, and linear equivalence property of the Weil's height machine (theorem 11.1).

Indeed, if we put $m = 1$ in (5), then we get the desired formula

$$[2]^*D \sim 3D + [-1]^*D.$$

For induction step, observe that

$$\frac{(m+1)^2+(m+1)}{2} = m^2 + m - \frac{(m-1)^2+(m-1)}{2} + 1,$$
$$\frac{(m+1)^2-(m+1)}{2} = m^2 - m - \frac{(m-1)^2-(m-1)}{2} + 1.$$

Assume (6). Then for $m+1$ we get

$$\begin{aligned} &\frac{(m+1)^2+(m+1)}{2} D + \frac{(m+1)^2-(m+1)}{2} [-1]^*D \\ &= 2\left(\frac{m^2+m}{2} D + \frac{m^2-m}{2} [-1]^*D\right) - \left(\frac{(m-1)^2+(m-1)}{2} D - \frac{(m-1)^2-(m-1)}{2} [-1]^*D\right) + D + [-1]^*D \\ &\sim 2[m]^*D - [m-1]^*D + D + [-1]^*D \\ &\sim [m+1]^*D \end{aligned}$$

Definition 12.5. A divisor D on an abelian variety A is called **symmetric** if $[-1]^*D \sim D$.

Corollary 12.6. *For a symmetric divisor D one has $[m]^*D \sim m^2 D$ and $h_{A,D}([m]P) = m^2 h_{A,D}(P) + O(1)$.*

Proof. This is immediate from the formula (6) and the Weil's height machine properties. □

Proposition 12.7 (Height parallelogram law)**.** *Let D be a symmetric divisor on an abelian variety A. Then for all $P, Q \in A(\overline{K})$*

$$h_{A,D}(P+Q) + h_{A,D}(P-Q) = 2h_{A,D}(P) + 2h_{A,D}(Q) + O(1).$$

This means that *up to some constant*, $h_{A,D}$ is a quadratic form.

Proof. Consider the following maps $A \times A \to A$:

$$\sigma(P,Q) := P+Q, \quad \delta(P,Q), \quad \pi_1(P,Q) := P, \quad \pi_2(P,Q) := Q.$$

By the "seesaw principle", we have an equivalence $\sigma^*D + \delta^*D \sim 2\,\pi_1^*D + 2\,\pi_2^*D$ on $A \times A$. Applying the Weil's height machine to this, we get

$$h_{A,\sigma^*D}(P,Q) + h_{A,\delta^*D}(P,Q) = 2\,h_{A,\pi_1^*D}(P,Q) + 2\,h_{A,\pi_2^*D}(P,Q) + O(1).$$

And then by functoriality the desired formula follows. □

Theorem 12.8 (Néron–Tate). *Let V/K be a smooth projective variety defined over a number field. Let $\phi\colon V \to V$ be a morphism such that $\phi^*D \sim \alpha D$ with $\alpha > 1$ for some divisor D. Then there exists a unique height function $\widehat{h}_{V,\phi,D}\colon V(\overline{K}) \to \mathbb{R}$, called the **Néron–Tate height** (or the **canonical height**) on V, with the following properties:*

(1) $\widehat{h}_{V,\phi,D}(P) = \lim_{n\to\infty} \alpha^{-n}\, h_{V,D}(\phi^n(P))$, *where* $\phi^n := \phi \circ \cdots \circ \phi$.

(2) $\widehat{h}_{V,\phi,D}(P) = h_{V,D}(P) + O(1)$.

(3) $\widehat{h}_{V,\phi,D}(\phi(P)) = \alpha\, \widehat{h}_{V,\phi,D}(P)$.

Note that in the equations (1) and (3) there is no bounded function "$O(1)$"—these are true equalities.

As for the condition $\phi^*D \sim \alpha D$, by the corollary 12.6, for an abelian variety $V = A$ we can take D to be a symmetric divisor and ϕ a multiplication by m map $[m]\colon A \to A$, where $m = 2,3,4,\ldots$

Proof. The property (1) tells that we need to consider the sequence $\alpha^{-n}\, h_{V,D}(\phi^n(P))$ for $n = 0,1,2,\ldots$ We show that it is Cauchy, and hence converges to some $\widehat{h}_{V,\phi,D}(P) := \lim_{n\to\infty} \alpha^{-n}\, h_{V,D}(\phi^n(P))$.

If we apply ϕ only once, then we have

$$h_{V,D}(\phi(P)) = \alpha\, h_{V,D}(P) + O(1),$$

so $h_{V,D}(\phi(P)) - \alpha\, h_{V,D}(P)$ is bounded:

$$|h_{V,D}(\phi(P)) - \alpha\, h_{V,D}(P)| \le C \quad \text{for all } P.$$

For the difference $\alpha^{-n}\, h_{V,D}(\phi^n(P)) - \alpha^{-m}\, h_{V,D}(\phi^m(P))$ we have

$$\begin{aligned} |\alpha^{-n}\, h_{V,D}(\phi^n(P)) - \alpha^{-m}\, h_{V,D}(\phi^m(P))| &= \left| \sum_{m+1\le i\le n} \alpha^{-i}\, h_{V,D}(\phi^i(P)) - \alpha^{-(i-1)}\, h_{V,D}(\phi^{i-1}(P)) \right| \\ &\le \sum_{m+1\le i\le n} \alpha^{-i} \cdot \left| h_{V,D}(\phi^i(P)) - \alpha\, h_{V,D}(\phi^{i-1}(P)) \right| \\ &\le \sum_{m+1\le i\le n} \alpha^{-i}\, C \\ &\le \frac{\alpha^{-m} - \alpha^{-n}}{\alpha - 1}\, C, \end{aligned}$$

and we conclude that the sequence is Cauchy.

Observe that as $n \to \infty$, we get

$$\left| \widehat{h}_{V,\phi,D}(P) - h_{V,D}(P) \right| \le \frac{C}{\alpha - 1},$$

hence the property (2).

For the property (3), observe that

$$\widehat{h}_{V,\phi,D}(\phi(P)) = \lim_{n\to\infty} \alpha^{-n} h_{V,P}(\phi^n(\phi(P))) = \lim_{n\to\infty} \alpha\, \frac{h_{V,D}(\phi^{n+1}(P))}{\alpha^{n+1}} = \alpha\, \widehat{h}_{V,\phi,D}(P).$$ □

On an abelian variety, from the height parallelogram law for $h_{A,D}$ (proposition 12.7), we have

$$\alpha^{-n} h_{A,D}(P+Q) + \alpha^{-n} h_{A,D}(P-Q) = \alpha^{-n} 2 h_{A,D}(P) + \alpha^{-n} 2 h_{A,D}(Q) + \alpha^{-n} O(1).$$

Taking the limit $n \to \infty$, we obtain the following.

Proposition 12.9. *Let A be an abelian variety over a number field K. Let D be a symmetric divisor on A such that $\phi^* D \sim \alpha D$ for some $\alpha > 1$. Then for all $P, Q \in A(\overline{K})$*

$$\widehat{h}_{A,\phi,D}(P+Q) + \widehat{h}_{A,\phi,D}(P-Q) = 2\widehat{h}_{A,\phi,D}(P) + 2\widehat{h}_{A,\phi,D}(Q).$$

So $\widehat{h}_{A,\phi,D}$ is a true quadratic form on the abelian group $A(\overline{K})$ (any function $h: A \to \mathbb{R}$ satisfying the parallelogram law is a quadratic form).

We would like to see when the quadratic form is nonnegative definite, i.e. when $\widehat{h}_{A,\phi,D}(P) \geq 0$ for all $P \in A(\overline{K})$, and characterize the points where $\widehat{h}_{A,\phi,D}(P) = 0$. For this we take D to be an ample divisor.

Theorem 12.10. *Let A be an abelian variety over a number field K.*

As before, let $\phi: A \to A$ and D be such that $\phi^* D \sim \alpha D$ for $\alpha > 1$. Assume that D is an ample divisor. Then

(1) $\widehat{h}_{A,\phi,D}(P) \geq 0$.

(2) $\widehat{h}_{A,\phi,D}(P) = 0$ *iff ϕ is* ***preperiodic*** *at P, meaning that the set*

$$\{P, \phi(P), \phi^2(P), \ldots\}$$

is finite (so the values $\phi^n(P)$ are eventually periodic).

(3) The set

$$\{P \in A(K) \mid \phi \text{ is preperiodic at } P\}$$

is finite.

Proof. For (1), observe that $h_{A,D}(Q) \geq O(1)$ as D is ample, and so $\alpha^{-n} h_{A,D}(\phi^n(P)) \geq \alpha^{-n} O(1)$. Taking the limit $n \to \infty$, we have that $\widehat{h}_{A,\phi,D} \geq 0$.

For (2), if ϕ is preperiodic at P, then the set $\{\phi^n(P)\}$ is repeating, and so is the set of heights $\{h_{A,D}(\phi^n(P))\}$. The height $h_{A,D}(\phi^n(P))$ is bounded, and $\widehat{h}_{A,\phi,D}(P) = \lim_{n\to\infty} \alpha^{-n} h(\phi^n(P)) = 0$.

For the converse, if P is a point such that $\widehat{h}_{A,\phi,D}(P) = 0$, then $\widehat{h}(\phi^n(P)) = \alpha^n \widehat{h}_{A,\phi,D}(P) = 0$ for all n. But now the points from the set

$$\{P, \phi(P), \phi^2(P), \ldots\}$$

should have bounded height $h_{A,D}(\phi^n(P))$, because $\widehat{h}_{A,\phi,D}(Q) = h_{A,D}(Q) + O(1)$. Thus the set $\{\phi^n(P)\}$ is finite by the Northcott's property.

Finally, we see that (3) is the Northcott's property. □

If we apply the theorem to the multiplication by m morphism $[m]: A \to A$ with $m = 2, 3, 4, \ldots$, then $[m]$ is preperiodic at P means that $[m]^k P = [m]^\ell P$ for some $k > \ell \geq 1$, i.e. that P is a torsion point. So we have the following.

Proposition 12.11. *If $\phi = [m]$ for any $m = 2, 3, 4, \ldots$, then $\widehat{h}_{A,\phi,D}(P) = 0$ iff P is a torsion point on $A(K)$. There are finitely many such points.*

Further, note that if $P \in A(K)$ and $Q \in A(K)_{\text{tors}}$, then $\widehat{h}_{A,\phi,D}(P) = \widehat{h}_{A,\phi,D}(P+Q)$, i.e. adding a torsion point does not change the Néron–Tate height. Indeed, if $[n]Q = 0$, then

$$\widehat{h}_{A,\phi,D}(P+Q) = \frac{1}{n^2} \widehat{h}_{A,\phi,D}([n](P+Q)) = \frac{1}{n^2} \widehat{h}_{A,\phi,D}([n]P) = \widehat{h}_{A,\phi,D}(P).$$

13 Mordell–Weil theorem

We know that there are finitely many torsion points $A(K)_{\text{tors}}$, and now our interest turns to the points of $A(K)$ that are non-torsion. It is very difficult to construct them explicitly—how would you pick a rational point $P \in A(K)$? However, there is a very strong finiteness result.

Theorem 13.1 (Mordell–Weil). *Let A be an abelian variety defined over a number field K. Then the abelian group $A(K)$ is finitely generated, i.e. there exist some points $P_1, \dots, P_r \in A(K)$ such that*

$$A(K) \simeq \mathbb{Z}P_1 \oplus \cdots \oplus \mathbb{Z}P_r \oplus A(K)_{\text{tors}}.$$

The number r is called the **rank** of A. It depends on the field K, and usually it grows with degree of K. It can also be zero (for small number fields).

Remark 13.2. The rank is interpreted by the **Birch–Swinnerton-Dyer conjecture** which says it equals the analytic rank $\operatorname{ord}_{s=1} L(A, s)$.

As for the finite part, it is very hard to bound the size of $A(K)_{\text{tors}}$ and even harder to find its generators.

We also underline that the theorem is about K-rational points, of course it is not true that e.g. $A(\mathbb{C}) \simeq \mathbb{C}^g/\Lambda$ is finitely generated.

First we assume the following.

Theorem 13.3 (Weak Mordell–Weil). *$A(K)/nA(K)$ is finite for any $n = 1, 2, 3, \dots$*

Actually, for us any $n \geq 2$ will do, but the theorem is true for all n. From this we will deduce the Mordell–Weil theorem, and then we sketch the proof of the weak Mordell–Weil.

We are going to use the Néron–Tate height. We pick a symmetric ample divisor D on A and build from it the function $\widehat{h}(P) := \widehat{h}_{A,\phi,D}(P)$, where $\phi = [m]$ for any $m \geq 2$. This $\widehat{h}$ is a positive definite quadratic form (vanishing on the finitely many torsion points), and so it makes sense to define $\|P\| := \sqrt{\widehat{h}(P)}$. Observe that $\|[n]P\| = n\,\|P\|$ for all $n = 0, 1, 2, 3, \dots$

Let $C_0 := \max_{1 \leq i \leq s} \|P_i\|$, where $\{P_1, \dots, P_s\} = A(K)/nA(K)$. Consider the set $\mathcal{S} := \{P \in A(K) \mid \|P\| \leq C_0\}$. This set is non-empty, it contains all points $P_1, \dots, P_s$ and all torsion points $A(K)_{\text{tors}}$. By the Northcott's property $\mathcal{S}$ is finite. We will show that $\mathcal{S}$ generates $A(K)$.

Consider a point $Q_0 \in A(K)$ such that $Q_0 \notin \mathcal{S}$. Then $\|Q_0\| > C_0$. Looking at the image of Q_0 in $A(K)/nA(K)$, we get $Q_0 = P_i + [n]\,Q_1$ for some $Q_1 \in A(K)$. Further,

$$\|[n]\,Q_1\| = n\,\|Q_1\| = \|Q_0 - P_i\| \leq \|Q_0\| + \|P_i\| \leq \|Q_0\| + C_0.$$

So we have $n\,\|Q_1\| \leq \|Q_0\| + C_0 < \|Q_0\| + \|Q_0\|$, the last inequality being strict, and $\|Q_1\| < \frac{2}{n}\,\|Q_0\| \leq \|Q_0\|$, since $n \geq 2$.

If $Q_1 \in \mathcal{S}$, then we are done. Otherwise, we produce similarly a point Q_2 with $\|Q_2\| < \|Q_1\| < \|Q_0\|$. Such a sequence of points with decreasing norm cannot be infinite thanks to the Northcott's property. For some i we should have $Q_i \in \mathcal{S}$. □

The argument as above is called a **proof by infinite descent** and goes back to Fermat.

As for the weak Mordell–Weil, we just state the arithmetic results that imply the theorem.

Fact 13.4. *If K'/K is an extension of number fields, then the map*

$$A(K)/nA(K) \to A(K')/nA(K')$$

has finite kernel.

Fact 13.5. *The multiplication by n map $[n]\colon A(\overline{K}) \to A(\overline{K})$ is surjective, with finite kernel, and*

$$\ker[n] = A[n] \simeq (\mathbb{Z}/n\mathbb{Z})^{2g}.$$

For $x \in A(K)$ and $\sigma \in \mathrm{Gal}(\overline{K}/K)$ pick $y \in A(\overline{K})$ such that $[n]y = x$. We let $t(\sigma, x) := \sigma(y) - y$. The latter lies in $A[n]$:

$$[n](\sigma(y) - y) = \sigma([n](y)) - [n]y = \sigma(x) - x = 0.$$

This map $t\colon \mathrm{Gal}(\overline{K}/K) \times A(K) \to A[n]$ is called **Kummer pairing**.

From now on assume that $A[n] \subset A(K)$ and $\mu_n \subset K$ (so that Kummer's theory works).

Fact 13.6. *The function $t(\sigma, x)$ is bilinear and it is well-defined (does not depend on the choice of y).*

Consider the field $L := K(y_1, y_2, y_3, \ldots)$ obtained by adding the coordinates of all points y_i with $[n]y_i \in A(K)$ (this might be an infinite extension).

Then t induces a nondegenerate pairing $t\colon \mathrm{Gal}(L/K) \times A(K)/nA(K) \to A[n]$. Consequently, $A(K)/nA(K)$ is finite iff L/K is a finite extension.

Fact 13.7. *Let $x \in A(K)$. Consider the field $K(\frac{1}{n}x)$ generated by all coordinates $y_1, \ldots, y_k$ of points y such that $[n]y = x$. Then the extension $K(\frac{1}{n}x)$ is Galois over K and the group $\mathrm{Gal}(K(\frac{1}{n}x)/K)$ is isomorphic to a subgroup of $A[n]$.*

Fact 13.8. *Let v be a place of good reduction for A. Then, if $v \nmid n$, then $A[n](K) \hookrightarrow \widetilde{A}(K(v))$, where $K(v)$ denotes the residue field of v.*

That is, the reduction map is injective on $A[n]$ if $v \nmid n$.

Fact 13.9. *The set*

$$\mathcal{S} := \{v \mid v \text{ is a bad reduction place for } A, \text{ or } v \mid n\}$$

is finite. The extension $K(\frac{1}{n}x)/K$ is unramified outside of $\mathcal{S}$.

Fact 13.10 (Hermite's theorem)**.** *For a number field K, an integer d, and a finite set of places $\mathcal{S}$ there are finitely many extensions of K of degree $\le d$ that are unramified outside of $\mathcal{S}$.*

(One bounds the degree and the discriminant, hence there are finitely many number fields.)

So if we take L to be the compositum of $K(\frac{1}{n}x)$, it must be a finite extension.

All the above imply the weak Mordell–Weil. □

Proofs of the listed facts can be found in [Hindry–Silverman, Chapter C].

Remark 13.11. We have a method of computing the rank of A. Since $A(K) \simeq \mathbb{Z}^r \oplus A(K)_{\mathrm{tors}}$, we have $A(K)/nA(K) \simeq (\mathbb{Z}/n\mathbb{Z})^r \oplus (A(K)_{\mathrm{tors}}/nA(K)_{\mathrm{tors}})$. So the cardinality of these sets is $x = n^r y$. In some cases, e.g. for $n = 2$, the calculations are easy. It is called the n-**descent**. See Silverman, *The Arithmetic of Elliptic Curves*, Chapter X.

14 Mordell conjecture

We will be interested in curves. A **curve** is a projective variety of dimension 1. The space of differentials on C has finite dimension, and this dimension is called the **genus** of C.

Example 14.1. An elliptic curve is given by an equation $Y^2Z = X^3 + aXZ^2 + bZ^3$. Over complex numbers we have an isomorphism from $\mathbb{C}/\Lambda$ given by $z \mapsto (\wp(z) : \wp'(z) : 1)$. The holomorphic differential is $\omega = \frac{d\wp(z)}{\wp'(z)} = dz$.

Example 14.2. If a plane curve is given by an affine equation $Y^2 = F(X)$ with $\deg F = 2g + 1$ or $2g + 2$, then C has genus g. Such a curve is called **hyperelliptic**. (As usual we understand by C the smooth projective curve having an affine model $Y^2 = F(X)$; further we assume that F has no multiple roots.)

Example 14.3. If a curve is smooth and it has degree d, then $g = \frac{(d-1)(d-2)}{2}$.

It is natural to ask whether curves of higher genus can carry a group law as elliptic curves. The answer is *no*: C is an abelian variety only when C has genus one and a rational point. However, a curve C of genus g always embeds in its **Jacobian** $\mathrm{Jac}(C)$, which is an abelian variety of dimension g.

Let K be a number field. Let C be a curve of genus g defined over K.

- If $g = 0$, then it is a conic. There can be either no rational points at all (e.g. for $X^2 + Y^2 = -1$ over $\mathbb{Q}$), or infinitely many of them. For instance, if $C/\mathbb{Q}$ has one rational point, then it is isomorphic to $\mathbb{P}^1$.
- If $g = 1$, then $C = E$ is an elliptic curve, provided $C(K) \neq \varnothing$, and $E(K) \simeq \mathbb{Z}^r \oplus E(K)_{\mathrm{tors}}$ by the Mordell–Weil theorem.
- If $g > 1$, then Mordell conjectured in 1922 that $C(K)$ is finite.

It was proved by Faltings in 1983 that the Mordell conjecture is true. But first we will examine an alternative proof using the techniques of Vojta (1987), as presented by Bombieri (1990).

We have an embedding $j\colon C \to \mathrm{Jac}(C) = J$ into the Jacobian. On J the **theta divisor** $\Theta := \underbrace{j(C) + \cdots + j(C)}_{g-1 \text{ times}}$ is ample and symmetric, hence we can use Θ to construct the Néron–Tate height $\widehat{h}_{J,\Theta}$ on the Jacobian. Since $\widehat{h}_{J,\Theta}$ is a positive definite quadratic form on $J(K) \setminus J(K)_{\mathrm{tors}}$, we can define a scalar product $\langle \cdot, \cdot \rangle$ on $J(K) \otimes \mathbb{R}$ via

$$\langle P, Q \rangle := \frac{1}{2} \left(\widehat{h}_{J,\Theta}(P+Q) - \widehat{h}_{J,\Theta}(P) - \widehat{h}_{J,\Theta}(Q) \right) \quad \text{for } P, Q \in J(K).$$

The set of rational points $C(K)$ is a subset of $J(K)$, and the map $J(K) \to J(K) \otimes \mathbb{R}$ has the torsion subgroup $J(K)_{\mathrm{tors}}$ as its kernel. We already know by the Mordell–Weil theorem that $J(K)_{\mathrm{tors}}$ is finite, hence it is sufficient to show that $(J(K) \otimes \mathbb{R}) \cap j(C)$ is a finite set.

The difficult point is the **Vojta's inequality**.

Theorem 14.4. *There exist two constants $\kappa_1 = \kappa_1(C)$ and $\kappa_2 = \kappa_2(g)$ such that for all $z \in C(\overline{K})$ one has*

$$\left\{ \begin{array}{c} \|z\| \geq \kappa_1 \\ \text{and} \\ \|w\| \geq \kappa_2 \|z\| \end{array} \right\} \Rightarrow \langle z, w \rangle \leq \frac{3}{4} \|z\| \cdot \|w\|.$$

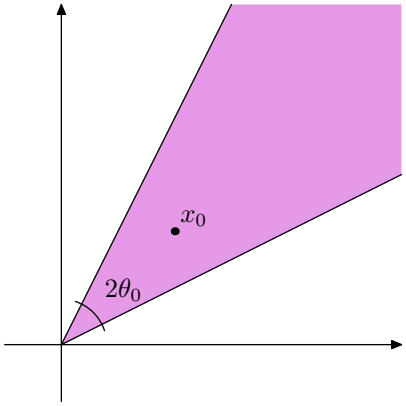
We note that $\langle z, w \rangle \leq \|z\| \cdot \|w\|$ by the Cauchy–Schwarz inequality, and Vojta's inequality is much stronger.

For two points $x, y \in J(K) \otimes \mathbb{R}$ we define the "angle" between them by

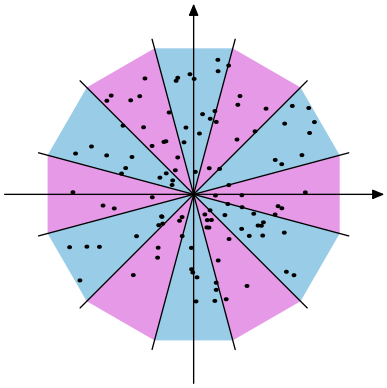
$$\cos \theta(x, y) := \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}.$$

For any point x_0 and any angle θ_0 we consider the cone Γ_{x_0,θ_0} consisting of points with angle less than θ_0:

$$\Gamma_{x_0,\theta_0} := \{ x \in J(K) \otimes \mathbb{R} \mid \theta(x, x_0) < \theta_0 \}.$$



It is clear that we can cover the whole space $J(K) \otimes \mathbb{R}$ by finitely many such cones with, say $\theta_0 = \frac{\pi}{12}$ (which is just an angle small enough to apply the Vojta's inequality). Inside each such cone consider the set $\Gamma_{x_0,\theta_0} \cap C(K)$. We need to conclude that it is finite, and then we are done.



For the sake of contradiction assume that $\Gamma_{x_0,\theta_0} \cap C(K)$ is infinite. It is possible to find $z \in \Gamma_{x_0,\theta_0} \cap C(K)$ with $\|z\| \geq \kappa_1$, otherwise $\Gamma_{x_0,\theta_0} \cap C(K)$ would be finite by the Northcott's property. Similarly, there exists a point $w \in \Gamma_{x_0,\theta_0} \cap C(K)$ such that $\|w\| \geq \kappa_2 \|z\|$. By the Vojta's inequality,

$$\langle z, w \rangle \leq \frac{3}{4} \|z\| \cdot \|w\|.$$

This means that $\cos\theta(z,w) \leq \frac{3}{4}$, but then $\theta(z,w) > \frac{\pi}{6}$. We took $\theta_0 = \frac{\pi}{12}$, and so this is not possible. This gives a contradiction and shows that each cone Γ_{x_0,θ_0} intersects $C(K)$ by finitely many points. □

☹☹☹ The essential difficulty is the Vojta's inequality. It is proved in part E of [Hindry–Silverman].

15 Some ingredients of the Faltings' proof

We just saw how the Vojta's inequality implies the Mordell conjecture. The original proof by Faltings was very different, and now we outline how it goes.

Shafarevič conjecture implies Mordell conjecture

An important finiteness result is the following.

Theorem 15.1 (Shafarevič conjecture). *Let K be a number field and let S be a finite set of primes in K. There are finitely many isomorphism classes of curves C/K of genus g with good reduction outside of S.*

This is an analogue of Hermite's theorem (fact 13.10). In fact Shafarevič conjecture implies the Mordell conjecture. This was showed by Parshin.

Theorem 15.2 (Kodaira–Parshin construction). *Let C/K be a curve of genus ≥ 2. If $C(K) \neq \varnothing$, then for a point $P \in C(K)$ there exists a curve C_P/K' and a morphism*

$$\phi_P \colon C_P \to C_{K'}$$

such that

1. *K'/K is a finite extension.*
2. *The genus $g(C_P)$ is bounded in terms of $g(C)$.*
3. *C_P has good reduction outside of a finite set of primes in K'.*
4. *ϕ_P is ramified exactly at P.*

This means that one can count curves C_P instead of rational points $P \in C(K)$, since to each pair (C_P, ϕ_P) corresponds exactly one point $P \in C(K)$. There can be various morphisms ϕ_P; however, there are finitely many of them due to the following classical result.

Theorem 15.3 (De Franchis). *Let C/K and C'/K be two curves over a field K. If $g(C) \geq 2$, then there are finitely many nonconstant maps $C' \to C$.*

Hence in the correspondence $P \mapsto C_P$ each curve C_P comes from finitely many points $P \in C(K)$. This means that Shafarevič conjecture implies the finiteness of $C(K)$.

Remark 15.4. De Franchis theorem is the only point that uses the assumption $g(C) \geq 2$.

In his paper Faltings proved the Shafarevič conjecture.

A finiteness theorem for abelian varieties

The Shafarevič conjecture can be deduced from the following theorem.

Theorem 15.5. *Fix a number field K, a finite set S of primes in K, and a number g.*
There are finitely many isomorphism classes of abelian varieties A/K of dimension g having good reduction at all primes outside S.

If a curve C has a good reduction at a prime $\mathfrak{p} \notin S$, then also does the Jacobian variety $\mathrm{Jac}(C)$. Further, we have the following.

Theorem 15.6 (Torelli). *An isomorphism class of a curve C/K is uniquely determined by the isomorphism class of the principally polarized abelian variety $(\mathrm{Jac}(C), \Theta)$.*

Hence to prove the Shafarevič conjecture, it is enough to prove theorem 15.5. This means we need to count *abelian varieties*.

Faltings defined height of an abelian variety $h_F(A/K)$ and proved the "Northcott property" for it: bounding the height h_F and dimension g gives finitely many isomorphism classes of varieties.

Faltings height of an abelian variety

Definition 15.7. Let A/K be an abelian variety of genus g. Then A extends canonically to a smooth group scheme $\mathcal{A}/\operatorname{Spec}\mathcal{O}_K$, the Néron model. Let $\epsilon\colon \operatorname{Spec}\mathcal{O}_K \to \mathcal{A}$ be the natural section whose image in each fiber is the zero element. We put

$$\omega_{\mathcal{A}/\operatorname{Spec}\mathcal{O}_K} := \epsilon^*(\Omega^g_{\mathcal{A}/\operatorname{Spec}\mathcal{O}_K}).$$

It is a projective $\mathcal{O}_K$-module of rank 1.

For each $s \in \omega_{\mathcal{A}/\operatorname{Spec}\mathcal{O}_K} \smallsetminus \{0\}$ the set $\omega_{\mathcal{A}/\operatorname{Spec}\mathcal{O}_K}/s\mathcal{O}_K$ is finite.

If v is an infinite prime in K, we set

$$\|s\|_v^2 := \frac{i^{g^2}}{c_0^g} \int_{A(\mathbb{C})} s \wedge \overline{s}.$$

Here c_0 is a constant, which is different for different authors (Faltings: $c_0 = 2$, Deligne: $c_0 = 2\pi$, Silverman: $c_0 = (2\pi)^2$, etc.).

The **differential height** (**Faltings height**) of A is given by

$$h_F(A/K) := \frac{1}{[K:\mathbb{Q}]}\left(\log \#(\omega_{\mathcal{A}/\operatorname{Spec}\mathcal{O}_K}/s\mathcal{O}_K) - \sum_{v\in M_K^\infty} \log \|s\|_v^{d_v}\right).$$

As always $d_v := [K_v : \mathbb{Q}_v]$. The quantity $h_F(A/K)$ does not depend on the choice of $s \neq 0$.

With $c_0 = (2\pi)^2$ in the definition above, one has $h_F(A/K) \geq 0$.

Remark 15.8. For elliptic curves it is relatively easy to write down the Faltings height.

With particular choice of s, we get

$$h_F(E/K) = \frac{1}{12\cdot[K:\mathbb{Q}]}\left(\log N_{K/\mathbb{Q}}|\Delta_E| - \sum_{v\in M_K^\infty} \log((2\pi)^{12}\cdot|\Delta(\tau_v)|\cdot \operatorname{Im}\tau_v)^{d_v}\right).$$

For details see *Arithmetic Geometry* (edited by Gary Cornell, Joseph H. Silverman), 1986, Chapter X, "Heights and elliptic curves".

Here Δ_E is the minimal discriminant of E, the numbers τ_v come from the uniformization $E(\overline{K}_v) = \mathbb{C}/\mathbb{Z} + \tau_v\mathbb{Z}$, and $\Delta(\tau_v)$ is the modular discriminant

$$\Delta(\tau) := q\prod_{n\geq 1}(1-q^n)^{24}, \quad q := e^{2\pi i\tau}.$$

In terms of $q := e^{2\pi i\tau}$ the j-invariant is given by a Laurent series

$$j(\tau) = \frac{1}{q} + 744 + 196884\,q + 21493760\,q^2 + 864299970\,q^3 + \cdots$$

Note that $|q| = e^{-2\pi \operatorname{Im}\tau}$. For $\operatorname{Im}\tau \gg 1$ one has $|j(\tau)| \approx e^{2\pi\operatorname{Im}\tau}$, and so

$$\log|j(\tau)| \approx 2\pi\operatorname{Im}\tau \approx -\log|q| \approx -\log|\Delta(\tau)| \approx -\log|\Delta(\tau)\ \operatorname{Im}\tau|.$$

Thus

$$h_F(E/K) \ll\gg \max\{\log N_{K/\mathbb{Q}}|\Delta_E|, h(j_E)\},$$

where $A \gg B$ denotes $A \geq C_1 B + C_2$ for some constants C_1 and C_2.

Indeed, for an elliptic curve E/K one can think of two fundamental invariants: the minimal discriminant and the j-invariant. It turns out they are related to Faltings height of E.

Theorem 15.9 (Faltings). *The differential height satisfies the Northcott's property. There are finitely many isomorphism classes of abelian varieties A/K of fixed dimension g and bounded Faltings height $h_F(A/K) \le M$.*

Using the Faltings height, one can prove the finiteness theorem for abelian varieties.

☹☹☹ The proof is difficult, and over the course one also proves the Tate conjecture. A good exposition of this can be found in http://jmilne.org/math/CourseNotes/av.html

16 Bounding the number of points

If C/K is a curve of genus $g \ge 2$, then we know by the Faltings' theorem that $C(K)$ has finitely many points. In fact it is possible to extract explicit bounds (in terms of g and arithmetic invariants of K) from the Vojta's counting arguments.

Theorem 16.1 (Rémond, 2000–2010).

$$\#C(K) \le \left(2^{38+2g} \cdot [K:\mathbb{Q}] \cdot g \cdot \max\{1, h_\Theta\}\right)^{(r_K+1)\,g^{20}}.$$

Here r_K is the rank of $\mathrm{Jac}(C)(K)$ (given by Mordell–Weil) and h_Θ is the height $h(\phi_\Theta(0))$, where $\phi_\Theta\colon \mathrm{Jac}(C) \hookrightarrow \mathbb{P}^N$ is the embedding given by theta-functions.

It is a conjecture of Lang that there is a constant $c(g,K)$ depending only on g and K such that

$$\#C(K) \le c(g,K)^{r_K+1}.$$

The result of Rémond is interesting from theoretical point of view but not for practical bounds. It is enough to mention that the current world's record for the number of rational points has the hyperelliptic curve of genus 2 given by (these are integer coefficients, not phone numbers!)

$$y^2 = 82342800\,x^6 - 470135160\,x^5 + 52485681\,x^4 + 2396040466\,x^3 + 567207969\,x^2 - 985905640\,x + 247747600.$$

It has 642 rational points (discovered by Michael Stoll using families constructed by Noam Elkies; see http://www.mathe2.uni-bayreuth.de/stoll/recordcurve.html). The Rémond's bound basically tells us that

$$642 \le (2^{43}\,h_\Theta)^{(r_K+1)\,2^{20}}.$$

17 The method of Chabauty and Coleman

There is another bound by Chabauty and Coleman, which can give optimal results in particular cases. For simplicity of presentation we work over $K = \mathbb{Q}$. The method can be applied if one knows the rank of $\mathrm{Jac}(C)(\mathbb{Q})$ by some kind of effective Mordell–Weil (for particular cases!).

Theorem 17.1 (Chabauty–Coleman). *Let C be a curve of genus ≥ 2 defined over $\mathbb{Q}$. Suppose that $\mathrm{rk}\,\mathrm{Jac}(C)(\mathbb{Q}) \le g-1$. Then one has*

$$\#C(\mathbb{Q}) \le \#C(\mathbb{F}_p) + 2g - 2$$

for any $p > 2g$ with C having good reduction at p.

The finiteness was first proved by Chabauty (1941), and Coleman (1985) obtained the bound.

Example 17.2 (Grant 1994). Consider a hyperelliptic curve C given by an affine equation

$$y^2 = x\,(x-1)\,(x-2)\,(x-5)\,(x-6).$$

The polynomial on the right hand side is of degree 5, so the genus is 2, and there are finitely many points $C(\mathbb{Q})$.

The rank of $\mathrm{Jac}(C)(\mathbb{Q})$ is 1, so we can apply Chabauty–Coleman. The prime $p = 7$ is of good reduction, and $7 > 2g$, so we know that $\#C(\mathbb{Q}) \le \#C(\mathbb{F}_7) + 2$.

It is not difficult to list the eight points over $\mathbb{F}_7$ satisfying the equation:

$$C(\mathbb{F}_7) = \{(0,0),\ (1,0),\ (2,0),\ (5,0),\ (6,0),\ (3,1),\ (3,6),\ (\infty,\infty)\}.$$

Further, it is not difficult to find ten points over $\mathbb{Q}$:

$$\{(0,0),\ (1,0),\ (2,0),\ (5,0),\ (6,0),\ (10,\pm 120),\ (3,\pm 6),\ (\infty,\infty)\}.$$

But the bound gives $\#C(\mathbb{Q}) \le \#C(\mathbb{F}_7) + 2 = 10$, so we just listed all rational points on C!

Of course in the last example we were very lucky. If the upper bound does not coincide with the number of discovered rational points, then we can't tell if there are more points, or it's just the bound is not sharp.

Example 17.3 (Cassels–Flynn). Consider a curve given by

$$y^2 = 2x\,(x^2 - 2x - 2)\,(-x^2 + 1).$$

It has genus 2. One can show by different methods that there are exactly six rational points:

$$(0,0),\ (\pm 1,0),\ \left(-\frac{1}{2}, \pm\frac{3}{4}\right),\ (\infty,\infty).$$

However, if $p \ge 5$, then $\#C(\mathbb{F}_p) \ge 5$, so Chabauty–Coleman gives $\#C(\mathbb{Q}) \le 7$.

Now we outline how the proof of the Chabauty–Coleman bound goes.

We consider the Jacobian $J = \mathrm{Jac}(C)$ with an embedding $C \hookrightarrow J$. With respect to this embedding, the rational points $C(\mathbb{Q})$ are contained in $J(\mathbb{Q})$. We consider the p-adic points $J(\mathbb{Q}_p)$, which is a p-adic Lie group, and then $C(\mathbb{Q})$ is contained in the set $C(\mathbb{Q}_p) \cap \overline{J(\mathbb{Q})}^p$, where $\overline{J(\mathbb{Q})}^p$ is the closure of $J(\mathbb{Q})$ in $J(\mathbb{Q}_p)$ with respect to the p-adic topology.

Chabauty showed that if we assume $\dim \overline{J(\mathbb{Q})}^p < g$, then the set $C(\mathbb{Q}_p) \cap \overline{J(\mathbb{Q})}^p$ is indeed finite. In particular, one has

$$\dim \overline{J(\mathbb{Q})}^p \le \mathrm{rk}\, J(\mathbb{Q}),$$

so it is enough to assume that $\mathrm{rk}\, J(\mathbb{Q}) \le g - 1$.

We have the $\mathbb{Q}_p$-vector space $H^0(J_{\mathbb{Q}_p}, \Omega^1)$ of regular 1-forms on $J_{\mathbb{Q}_p}$ (which is the variety obtained from J by extension of scalars). There is a bilinear pairing

$$\begin{aligned} J(\mathbb{Q}_p) \times H^0(J_{\mathbb{Q}_p}, \Omega^1) &\to \mathbb{Q}_p, \\ (Q, \omega_J) &\mapsto \int_O^Q \omega_J. \end{aligned}$$

The integral $\int_O^Q \omega_J$ is defined as the unique homomorphism $Q \mapsto \int_O^Q \omega_J$ which locally on an open subgroup $U \ni Q$ is computed by formal integration of a power series expansion of ω_J.

An embedding $C \hookrightarrow J$ induces the restriction map $H^0(J_{\mathbb{Q}_p}, \Omega^1) \to H^0(C_{\mathbb{Q}_p}, \Omega^1)$.

Coleman showed the following.

Fact 17.4 (Coleman 1985). *Assume that the curve $C/\mathbb{Q}$ has good reduction at p. Let $\omega \in H^0(C_{\mathbb{Q}_p}, \Omega^1)$ be a nonzero 1-form which is a restriction of $\omega_J \in H^0(J_{\mathbb{Q}_p}, \Omega^1)$. Scale ω by an element of $\mathbb{Q}_p^\times$, so that it reduces to a nonzero 1-form $\widetilde{\omega} \in H^0(C_{\mathbb{F}_p}, \Omega^1)$.*

Consider the reduction map $C(\mathbb{Q}_p) \twoheadrightarrow C(\mathbb{F}_p)$ (which is surjective). For a point $\widetilde{Q} \in C(\mathbb{F}_p)$ let $m_{\widetilde{Q}} := \operatorname{ord}_{\widetilde{Q}} \widetilde{\omega}$. If $m_{\widetilde{Q}} < p-2$, then the number of points $Q \in C(\mathbb{Q})$ reducing to $\widetilde{Q}$ is at most $m_{\widetilde{Q}} + 1$.

One has by the Riemann–Roch theorem

$$\sum_{\widetilde{Q} \in C(\mathbb{F}_p)} m_{\widetilde{Q}} \le 2g - 2,$$

so under assumption $p > 2g$ one gets $m_{\widetilde{Q}} \le 2g - 2 < p - 2$, and by the Coleman's result above

$$\#C(\mathbb{Q}) \le \sum_{\widetilde{Q} \in C(\mathbb{F}_p)} (m_{\widetilde{Q}} + 1) = \#C(\mathbb{F}_p) + \sum_{\widetilde{Q} \in C(\mathbb{F}_p)} m_{\widetilde{Q}} \le \#C(\mathbb{F}_p) + 2g - 2.$$

For more details see W. McCallum, B. Poonen, *The Method of Chabauty and Coleman.*

Of course bounds can be pretty sharp for specific curves, but in general bounding the number of points is very difficult. What is even more difficult is to bound the *height* of rational points. In theory we can enumerate all the rational points on $\mathbb{P}^N$ and check whether each of them lies on the curve. If we have just an upper bound on the number of points, then we do not know when to stop this enumeration: suppose we have found 642 rational points, how do we know that we should stop and there are no more of them? If the upper bound is not tight, then it is not possible. However, if we manage to bound the *height*, then we know precisely when to stop enumerating the points (by increasing height).

Our final question concerns bounding the height of points from below. It is also very hard, but there are plausible conjectures supported by partial results.

18 Bounding the height of points from below

On an abelian variety it is natural to ask for a uniform bound on the height $\widehat{h}(P)$ from below.

Conjecture 18.1 (Lang–Silverman, simplified statement). *Let A/K be an abelian variety over a number field. Assume that A is simple, i.e. has no proper abelian subvarieties. Let $g \ge 1$ be the dimension. Then there exist constants $c_1(g,K) > 0$ and $c_2(g,K) > 0$ such that*

(1) either P is a torsion point, and $[n]P = 0$ for $1 \le n \le c_1(g,K)$,

(2) or P is non-torsion, and $\widehat{h}_{A,D}(P) \ge c_2(g,K) \cdot \max\{1, h_F(A/K)\}$ for any ample symmetric divisor $D \in \operatorname{Div}(A)$.

In particular, it is conjectured that $\#A(K)_{\text{tors}} \le c_1(g,K)^{2g}$, where $c_1(g,K)$ is a universal constant depending only on g and K.

In particular cases, e.g. for elliptic and hyperelliptic curves, there are available results towards the Lang–Silverman conjecture.

David (1993) and Masser (1993) showed that there is a family of hyperelliptic curves with varying genus g such that Lang–Silverman conjecture is true for $\operatorname{Jac}(C_n)$.

Hindry and Silverman showed that the Lang–Silverman conjecture for elliptic curves boils down to a conjecture about Szpiro quotients.

For an elliptic curve we define the **Szpiro quotient** to be

$$\sigma_{E/K} := \frac{\log |N_{K/\mathbb{Q}}(\Delta_{E/K})|}{\log |N_{K/\mathbb{Q}}(\mathfrak{f}_{E/K})|},$$

where $\Delta_{E/K}$ is the discriminant and $\mathfrak{f}_{E/K}$ is the conductor.

Theorem 18.2 (Hindry–Silverman, 1988). *There exists a constant $C = C([K:\mathbb{Q}], \sigma(E/K)) > 0$ such that for all non-torsion points $P \in E(K)$*

$$\widehat{h}(P) \geq C \cdot \max\{h(j_E), \log N_{K/\mathbb{Q}}(\Delta_{E/K})\}.$$

The bound can be written more explicitly. For example,

Theorem 18.3 (Petsche, 2006). *For non-torsion points $P \in E(K)$ one has*

$$\widehat{h}(P) \geq (10^{15} d^3 \sigma_{E/K}^6 \log^2(104613 d \sigma_{E/K}^2))^{-1} \cdot \log |N_{K/\mathbb{Q}}(\Delta_{E/K})|.$$

And in fact $\sigma_{E/K}$ is conjectured to be bounded.

Conjecture 18.4 (Szpiro). *For any $\epsilon > 0$ there exists a constant $C = C(K, \epsilon)$, such that for any elliptic curve E/K holds*

$$\log |N_{K/\mathbb{Q}}(\Delta_{E/K})| \leq (6 + \epsilon) \log |N_{K/\mathbb{Q}}(\mathfrak{f}_{E/K})| + C.$$

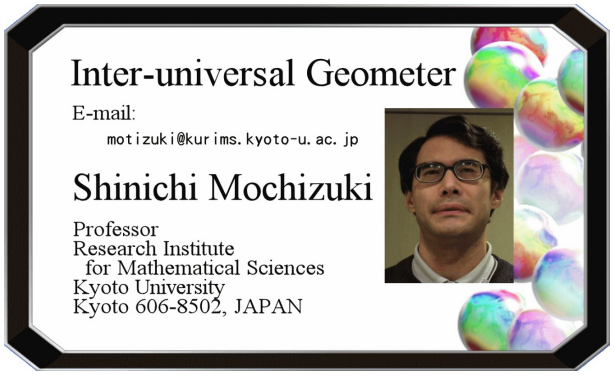
So with the Szpiro conjecture, Hindry–Silverman proves Lang–Silverman conjecture for elliptic curves. The Szpiro conjecture is more or less (up to changing constants) equivalent to the famous abc conjecture:

Conjecture 18.5 (Masser-Oesterlé). *Given $\epsilon > 0$ there exists $C(\epsilon)$ such that if $a, b, c \in \mathbb{Z}$ are nonzero and $a + b = c$ and $\gcd(a, b, c) = 1$, then*

$$\max\{|a|, |b|, |c|\} \leq C(\epsilon) \cdot \mathrm{rad}(abc)^{1+\epsilon},$$

where $\mathrm{rad}(abc) := \prod_{p | abc} p$.

A proof of abc, relying on something named "inter-universal Teichmüller theory", was claimed in 2012 by a Japanese mathematician Shinichi Mochizuki.



Here is another result, due to Fabien Pazuki. If A is an abelian variety of dimension 2, then it is a Jacobian of some curve, or a product of two elliptic curves. Namely,

Fact 18.6. *Let A/K be a principally polarized abelian variety of dimension 2. Then*

(1) either $A \simeq \mathrm{Jac}(C)$ for a curve C of genus 2, polarized by $\Theta = C$,

(2) or $A \simeq E_1 \times E_2$ is a product of elliptic curves, polarized by $\Theta = E_1 \times \{O\} + \{O\} \times E_2$.

For an archimedian place v on K we have uniformization $A(\overline{K}_v) \simeq \mathbb{C}^2/\mathbb{Z}^2 + \tau_v \mathbb{Z}^2$, where $\tau_v = \begin{pmatrix} \tau_{1,v} & \tau_{12,v} \\ \tau_{12,v} & \tau_{2,v} \end{pmatrix}$. We call the **archimedian trace** of A the quantity

$$\mathrm{Tr}_\infty(A) := \sum_{v \in M_K^\infty} d_v \,\mathrm{Tr}(\mathrm{Im}\,\tau_v),$$

and the **archimedian simplicity** of A is the quantity

$$s_\infty(A) := \prod_{v \in M_K^\infty} |\tau_{12,v}|_v^{d_v}.$$

One has $s_\infty(A) = 0 \iff A \simeq E_1 \times E_2$.

Theorem 18.7 (Pazuki, 2012). *In case $s_\infty(A) \neq 0$, so that $A \simeq \mathrm{Jac}(C)$, for a point $P \in A(K)$*

(1) either $[n]P = 0$ for $1 \leq n \leq c_1(d)$,

(2) or $\widehat{h}_{A,2\Theta}(P) \geq c_2(d) \cdot \left(\mathrm{Tr}_\infty(A) - \frac{5}{3}\frac{N_{K/\mathbb{Q}}(D)}{s_\infty(A)}\right)$, where $D := 2^8 \,\mathrm{disc}(F)$, and F is an integral model of the hyperelliptic curve $C: y^2 = F(x)$.

As a corollary, in case $\mathrm{Tr}_\infty(A) > \frac{5}{3}\frac{N_{K/\mathbb{Q}}(D)}{s_\infty(A)}$ one obtains the Lang–Silverman conjecture.
The relation of Tr_∞ to the Faltings height is the following:

$$h_F(A/K) \leq c_3(d) \cdot \mathrm{Tr}_\infty(A) + c_4(d)\,\frac{N_{K/\mathbb{Q}}(D)}{s_\infty(A)},$$

for some $c_3(d) > 0$, $c_4(d) > 0$.
For details see http://arxiv.org/abs/0812.2854v2.

Recall the Lehmer's conjecture for numbers $h(\alpha) \overset{?}{\geq} \frac{1}{[\mathbb{Q}(\alpha):\mathbb{Q}]}\,C$ (where α is not zero and not a root of unity). Similarly we can ask whether for an abelian variety A/K there is a constant $C(A) > 0$ such that $\widehat{h}_A(P) \overset{?}{\geq} \frac{1}{[\mathbb{Q}(P):\mathbb{Q}]}\,C(A)$ (where P is not a torsion point). Here we fix A and consider varying K, while in Lang–Silverman we fix K and vary A. One result in this direction is the following.

Theorem 18.8 (Ratazzi, 2004). *Let E/K be an elliptic curve with complex multiplication. There exists a constant $c(E/K) > 0$ such that for all $P \in E(\overline{K}) \setminus E(\overline{K})_{\mathrm{tors}}$*

$$\widehat{h}_E(P) \geq \frac{c(E/K)}{D}\left(\frac{\log\log 5D}{\log 2D}\right)^{13},$$

where $D := [K^{\mathrm{ab}}(P) : K^{\mathrm{ab}}]$.

See http://arxiv.org/abs/math/0402225.

Remark 18.9. Recall that an elliptic curve E has **complex multiplication** if it has nontrivial endomorphisms, which means $\mathrm{End}(E) \supsetneq \mathbb{Z}$.
For example, the curve $E: y^2 = x^3 - x$ has an extra endomorphism given by $i: (x, y) \mapsto (-x, iy)$.
Over finite fields an elliptic curve always has extra endomorphisms coming from the Frobenius map $x \mapsto x^p$. However, over a number field the property of having extra endomorphisms is exceptional.

In higher dimensions, for abelian varieties with complex multiplication, there are similar results by María Carrizosa.