Nils-Peter Skoruppa

HEIGHTS

Notes d'un cours de DEA

Ecole doctorale de Mathématiques — Université Bordeaux 1

Version: Id: heights.tex,v 1.3 2003/11/20 13:06:09 fenrir Exp

Preface

These are the notes of a *cours de DEA avancé* held at Bordeaux in spring 1998. The aim of the course was to introduce the notion of height, one of the basic ingredients in Diophantine geometry, in an elementary and easy to understand manner, with the emphasis on results, open problems and 'highlights' instead of abstract theory.

Accordingly we start in Part 1 with the classical Lehmer conjecture and discuss the important theorems around and towards this conjecture. In particular, we discuss Langevin's theorem and Zhang's theorem. When I prepared the course, it came to my mind that the theorem of Langevin, the more recent theorem of Zhang and the long-known result of Schintzel on the absolute bound for heights of real-algebraic numbers seem to have some deep analogy. In the present notes I tried to work this out, and in the end I managed (at least) to give a sort of unified proof for these results.

In Part 2 we discuss, after a generalisation of Zhang's theorem to plane affine algebraic curves, heights on elliptic curves. We discuss in an explicit manner the method of infinite descent (and Mordell's theorem), and the local decomposition of the canonical height, i.e. the "local Green's functions" on an elliptic curve. In the Appendix (which is actually an examination given to the students at the end of the course) the reader finds a sketch of how to compute explicitly the canonical height function on an explicitly given cubic algebraic curve.

A logical step would have been a third part treating Green's functions an algebraic curves of arbitrary genus, and, in particular, to have very concrete examples, a part treating Green's functions on modular curves.

These notes are still preliminary: a section on elliptic curves is missing (section 2.4), more recent considerations on the relation of Mahler measures and special values of certain L-functions and Mahler measures as entropy of certain "algebraic dynamical systems" are missing. Also the list of references is not complete, and the Appendix is in French. Maybe we shall come back to this (and also the Part 3) at another occasion.

We finally note that parts of Part 1 are based on a course given by the author in the Max-Planck-Institute für Mathematik in spring 1993. Criticism, comments and pointers to typos are welcome.

Talence, March 8, 1999
Nils-Peter Skoruppa

Contents

Part 1

Heights of Algebraic Numbers

It is natural to try to associate to an algebraic solution of a Diophantine equation a measure of complexity. This is natural in view of the problem of computing and storing such a number, but it has also a theoretical significance if the measure of complexity can be chosen such that the number of measured objects in question below a given bound is always finite.

In this first part we shall consider the problem of finding such a measure for algebraic numbers. This will lead to the notion of height for numbers. We shall discuss various properties of the height function, and in particular we shall discuss the Lehmer conjecture.

1.1 The height of a rational number

Assume that $\alpha = \frac{x}{y}$ is a rational number, say $\gcd(x, y) = 1$. We define its height by

$$H(\alpha) := \max(|x|, |y|).$$

This clearly measures the complexity of α in the sense of how many information do we need to describe α. Indeed, $\log H(\alpha)$ is roughly the number of digits needed to write down the numerator or denominator of α. Moreover it is clear that the set

$$\{\alpha \in \mathbb{Q} : H(\alpha) < B\}$$

is finite for any real B.

There is one important property that one can already read off in this more or less trivial situation. The height function possesses a decomposition into local factors. We explain this in detail.

Recall that to each (rational) prime p we can associate the valuation $|\cdot|_p$ of $\mathbb{Q}$ defined by

$$|\alpha|_p = p^{-n},$$

where p^n is the exact power of p in the prime decomposition of α. A valuation of a field K is a function $v : K^* \to \mathbb{R}_{\geq 0}$ such that $v(\alpha) = 0$ if and only if $\alpha = 0$ and satisfying

$$v(\alpha\beta) = v(\alpha)v(\beta), \quad v(\alpha + \beta) \leq v(\alpha) + v(\beta)$$

for all $\alpha, \beta \in K^*$. Two valuations v and w are called equivalent if there is a real number $s > 0$ such that $v(\alpha) = w(\alpha)^s$ for all $\alpha \in K$. Any valuation of $\mathbb{Q}$ is either equivalent to a $|\cdot|_p$ or to the usual absolute value on $\mathbb{Q}$ which we denote by $|\cdot|_\infty$ [Neuk], p. 124. The latter writing suggest that the set of primes of $\mathbb{Q}$ should be completed by a "prime at infinity".

Theorem 1.1. *For each rational number $\alpha \neq 0$ one has*

$$H(\alpha) = \max(1, |\alpha|_\infty) \prod_p \max(1, |\alpha|_p).$$

Proof. As before let x and y denote the numerator and denominator of α. The contribution from the pth factor on the right equals p^{-n}, where p^n is the exact divisor of α, if n is negative, and it equals 1 otherwise. Thus the product over the primes equals $|y|$. The factor before the product is 1 if $|x| < |y|$, and it is $|x|/|y|$ otherwise. This proves the formula. □

In view of the theorem it is reasonable to call the function

$$H_p(\alpha) := \max(1, |\alpha|_p)$$

the local height of α at the prime p. The decomposition formula of H can then be rewritten in a more compact form as

$$H = \prod_p H_p,$$

where this time p runs through the finite primes and $p = \infty$.

1.2 The Mahler measure of a polynomial

An algebraic number α is, up to "equivalence", described by its unique normalised minimal polynomial f. By the last we understand the minimal polynomial whose coefficients are in $\mathbb{Z}$ and are relatively prime. Discussing the complexity of α is thus equivalent to discussing the complexity of f.

To measure the complexity of a polynomial

$$f = a_n X^n + a_{n-1} x^{n-1} + \cdots + a_0 \in \mathbb{Z}[x]$$

we may consider the number

$$\|f\|_1 := \sum_{j=0}^{n} |a_j|.$$

This is in essence the number of digits needed to write down f. However, one might find good arguments to consider

$$|f\|_\infty = \max_j |a_j|$$

or

$$|f\|_2 := \sqrt{a_d^2 + \cdots + a_0^2}$$

as complexity measure.

Obviously we would prefer a unique, canonical one instead of many. Now the above three examples all come from a norm on the real vector space $\mathbb{R}[x]_n$ of real polynomials of degree less or equal to n. All such norms are equivalent, i.e. if $\|\cdot\|$ is any norm on $\mathbb{R}[x]_n$, then there exist constants $A, B > 0$ such that

$$A\,\|f\|_\infty \leq \|f\| \leq B\,\|f\|_\infty$$

for all f (exercise). Suppose we could construct for any f in a canonical way a sequence of polynomials f_k of degree less or equal to n such that, for any norm $|\|\cdot\|$, the measures $\|f_k\|^{\frac{1}{k}}$ are roughly $\|f\|$, and such that $\|f_k\|^{\frac{1}{k}}$ converges. By the last property the limit would not depend on the special choice of the norm as follows easily from the equivalence inequalities (see the proof of the next lemma for details). The limit can thus be considered as a good candidate for a canonical measure of complexity.

Such a sequence f_k can indeed be constructed. Let, for the following, f denote a polynomial with complex coefficients, say

$$f(x) = a_n x^n \cdots + a_0 = a_n \prod_{j=1}^{n} (x - \alpha_j).$$

We define

$$f_k(x) = a_n^k \prod_{j=1}^{n} (x - \alpha_j^k) = (-1)^{k(n+1)} \prod_{\zeta^n = 1} f(x^{\frac{1}{k}} \zeta).$$

Here the product is over all kth roots of unity. One might think of $\|f_k\|^{1/k}$ as being obtained by a sort of averaging over the roots of f. Note that f_k has rational coefficients if f has rational coefficients (since f_k is invariant under the Galois group of the decomposition field of f), which are, moreover, integral if those of f are integral.

We define the Mahler measure of f by

$$\mu(f) = |a_n| \prod_{j=1}^{n} \max(1, |\alpha_j|).$$

Thus $\mu(f)$ is, up to the number $|a_n|$ the product of the roots of f outside the unit circle, where multiple roots are repeated. We shall need a formula expressing $\mu(f)$ without making explicit reference to the zeroes of f.

Theorem 1.2. *(Jensen's formula) For any $f \in \mathbb{C}[X]$, $f \neq 0$ one has*

$$\log \mu(f) = \int_0^{2\pi} \log |f(e^{it})| \, dt.$$

Proof. Since the logarithm is additive It suffices to consider the case $f(z) = z - \alpha$. If $|\alpha| > 1$ then $\log |f(z)|$ is a harmonic function in a neighbourhood of the unit circle, and hence the integral equals

$$\log |f(0)| = \log |\alpha|.$$

If $|\alpha| < 1$ then $g(z) = 1 - \overline{\alpha} z$ has no zeroes in the unit circle, and $\log |g(x)|$ is a harmonic function in a neighbourhood of the unit circle. Moreover, $|g(z)| = |f(z)|$ on the unit circle. The integral in question thus equals the same integral but with f replaced by g, i.e. it equals

$$\log |g(0)| = 0.$$

Finally, if $|\alpha| = 1$, then

$$\frac{1}{2\pi} \int_0^{2\pi} \log |\mathrm{e}^{it} - \alpha| \, dt = \frac{1}{2\pi} \int_0^{2\pi} \log |\mathrm{e}^{it} - 1| \, dt = 0$$

(exercise). □

The actual formula known as Jensen's formula in complex analysis applies to a slightly more general functions than only to polynomials as stated in the theorem [Ahlf], p. 205.

We need a second property of the mahler measure.

Theorem 1.3. *(Norm inequality) For all $0 \leq j < n$ one has*

$$|a_j| \leq \binom{n}{j} \mu(f).$$

Proof. This is an immediate consequence of

$$a_j = (-1)^{n-j} a_n \sum_{\{j_1,\dots,j_{n-j}\} \subset \{1,\dots,n\}} \alpha_{j_1} \cdots \alpha_{j_{n-j}}$$

and the very definition of the Mahler measure. □

We are now able to explain why the Mahler measure is the canonical complexity measure we are looking for.

Theorem 1.4. *Let $\|\cdot\|$ be a norm on the real vector space $\mathbb{C}[x]_n$. Then, for any polynomial $f \in \mathbb{C}[x]_n$, one has*

$$\lim_{k\to\infty} \|f_k\|^{1/k} = \mu(f).$$

Proof. From the equivalence inequality we obtain

$$A^{1/k} \|f_k\|_\infty^k \leq \|f_k\|^{1/k} \leq B^{1/k} \|f_k\|_\infty^{1/k}$$

Thus, if the theorem holds true for the $\|\cdot\|_\infty$-norm, then it holds true for all norms.

For the $\|\cdot\|_\infty$-norm we have

$$\mu(f) \leq (n+1)\,\|f\|_\infty \leq 2^n (n+1)\,\mu(f).$$

The first inequality follows from Jensen's formula for $\mu(f)$ on using

$$|f(x)| \leq (n+1)\,\max|a_j|$$

for $|x| = 1$. The second one is an easy consequence of the norm inequality. Now by the very definition of $\mu(f)$ one has

$$\mu(f) = \mu(f_k)^{1/k}.$$

Combined with the above inequalities this gives

$$\mu(f) \leq \left[(n+1)\,\|f_k\|_\infty\right]^{1/k} \leq [2^n(n+1)]^{1/k} \mu(f).$$

Letting k tend to infinity we recognise the asserted formula. □

We note that there are other possibilities for defining the complexity of a polynomial f over $\mathbb{Z}$. One might consider for example $|f(1)|$, i.e. the absolute value of the sum of the coefficients of f. Again $|f_k(1)|^{\frac{1}{k}} \to \mu(f)$: indeed $\frac{1}{k}\log|f_k(1)|$ is just the $n-th$ Riemann sum approximating the integral defining $\log\mu(f)$.

In the computer algebra system Pari [Pari] one finds the function "polred" which finds for a given unitary integral f of degree n a new polynomial g which defines the same number field but which is (probably) minimal with respect to the function

$$l(f) = \sqrt{\alpha_1^d + \cdots \alpha_n^2}$$

[CoDi]. It is easy to verify that Again $l(f_k)^{\frac{1}{k}}$ tends to $\mu(f)$ for any f.

We note two simple but remarkable properties of the Mahler measure.

Theorem 1.5. *Let f and g be any complex polynomials. Then*

$$\mu(fg) = \mu(f)\mu(g), \quad \mu(f^*) = \mu(f).$$

Here f^* is the reciprocal polynomial of f, i.e. $f^*(x) = x^{\deg f} f(1/x)$.

Proof. The first identity is evident from the definition of μ. The second one is equivalent to

$$\frac{|a_0|}{|a_n|} = \prod_{j=1}^{n} |\alpha_j|.$$

□

By the norm inequality we see that, for any degree n and any bound B, there are only finitely many polynomials $f \in \mathbb{Z}[x]_n$ such that $\mu(f) \leq B$. In particular, for any real A the number

$$\inf\{\mu(f) : f \in \mathbb{Z}[x]_n,\ \mu(f) > A\}$$

is strictly greater 0 and is attained by a finite number of $f \in \mathbb{Z}[x]^n$. Obviously the Mahler measure of an integral polynomial is always greater or equal to 1. Thus it is natural to ask first of all for those polynomials with Mahler measure 1. This question is answered by a classical theorem.

Theorem 1.6. *(Kronecker) Let $f \in \mathbb{Z}[X]$. Then $\mu(f) = 1$ if and only if all roots of f are roots of unity or* 0*.*

Proof. Assume that $in \mathbb{Z}[x]$ has degree n and Mahler measure 1. For the j-th coefficient $a_j^{(k)}$ of f_k we have by the norm estimate and since $\mu(f_k) = 1$ the estimate

$$|a_j^k| \leq \binom{n}{j}.$$

Thus the set of all f_k is actually finite. Hence, the set of all roots of all f_n is finite. In particular, if α is a root of f then the α_n cannot be pairwise different. Consequently $\alpha^k = \alpha^l$ for some $k > l$, i.e. either $\alpha = 0$ or else $\alpha^{k-l} = 1$.

The inverse direction of the theorem is trivial. □

The theorem is usually cited in the form that an integral algebraic number whose conjugates are less or equal to 1 is necessarily a root of unity. Note that the statement in this form becomes false if one drops the integrality assumption; counter example: $\frac{3+4i}{5}$.

In view of the preceding theorem one is naturally interested in the numbers

$$\inf\{\mu(f) : f \in \mathbb{Z}[X]_n, \mu(f) > 1\}$$

and the polynomials realizing these Mahler measures. For a given degree n these minimizing polynomials are easy to calculate. In fact, one simply lists all polynomials $f \in \mathbb{Z}[x]_n$ with, say, $\mu(f) \leq 2$. This list is not empty since $\mu(x-2) = 2$, and it is contained in the finite set S_n of all integral f with

$$|a_j| \leq 2\binom{d}{j}$$

by the norm inequality. Thus this list can be compiled by searching S_n. However note that e.g. for $n = 4$ the set S_4 comprises already

$$(4\binom{4}{0} + 1)^2(4\binom{4}{1} + 1)^2(4\binom{4}{2} + 1) = 180625$$

elements. This can of course be cut down by some factor on using $\mu(f^*) = \mu(f)$, $\pm\mu(f(\pm x)) = \mu(f)$ by rejecting all polynomials with leading and constant term different from ±1. In Table 1.2 we listed the result of such a computational research for degrees $n \leq 5$.

n	f	$\mu(f)$	$\mathrm{disc}(f)$
1	$x-2$	2	2
2	x^2-x-1	$1.618\ldots$	5
3	x^3-x+1	$1.324\ldots$	-23
4	x^4-x^3-1	$1.380\ldots$	-283
5	$x^5-x^4+x^3-x+1$	$1.349\ldots$	$17\cdot 97$

Table 1.1: This table gives, for a given degree $n \leq 5$, a polynomial f from the set T_n whose Mahler measure is minimal. Here T_n is the set of all irreducible polynomials in $\mathbb{Z}[x]$ of degree n with Mahler measure strictly greater than 1. The respective minimum is also attained by the polynomials $\pm f(\pm x)$ and $\pm x^n f(\pm 1/x)$, but by no other ones in T_n.

1933 Lehmer conjectured [Lehm] that even the number

$$\mu_1 := \{\mu(f) : f \in \mathbb{Z}[X],\ \mu(f) > 1\}$$

is strictly greater than one. He even conjectured that μ_∞ is assumed by the polynomial

$$f_L(x) = x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1.$$

Here

$$\mu(f_L) = 1.176\ldots$$

The conjecture is still unproven and Lehmer's lower bound is still not beaten. A huge amount of computations has been done [Boyd] giving evidence to Lehmer's conjecture.

1.3 Pisot and Salem numbers

In view of the Lehmer conjecture it is an amusing sport to find polynomials f in $\mathbb{Z}[x]$ with minimal Mahler measure $\mu(f)$. A first naive approach is to look at polynomials with small $\|f\|_\infty$ norm, say with coefficients equal to ± 1. Systematic searches in this direction have been done e.g. in [Boyd].

A more theoretic approach is to search for algebraic numbers who are not "too far away" from roots of unity. Indeed, since the Mahler measure is multiplicative and greater and equal to 1 it suffices to look at irreducible polynomials f. Moreover, since $\mu(f)$ is greater than or equal to the constant and the leading term of f, it suffices to look at polynomials where both are equal to 1, i.e. at minimal polynomials f of algebraic units α. Now, one might expect that the Mahler measure $\mu(f)$ is small if many of the conjugates of α lie in the unit disk or on the unit circle.

An integral algebraic number α is called a Pisot number if $\alpha > 1$ and all its conjugates α' satisfy $|\alpha'| < 1$. It is called a Salem number if $\alpha > 1$, if all its conjugates α' satisfy $|\alpha'| \leq 1$, but if at least one α' satisfies $|\alpha'| = 1$.

A Salem number satisfies actually a stricter condition.

Theorem 1.7. *Let f be the normalized minimal polynomial of a Salem number τ. Then $f^* = f$.*

Proof. Let τ' be a conjugate of τ on the unit circle. Then τ' is a root of f and of f^*. Thus $f^* = \pm f$. If $f^* = -f$ then $f(1) = 0$, which is impossible. □

Hence, for a Salem number τ, the set of its conjugates is stable under $z \mapsto 1/z$. In particular, a Salem number is a unit, and moreover we have:

Corollary 1.7.1. *An algebraic integer τ is a Salem number if and only if $1/\tau$ is conjugate to τ and all other conjugates of τ have absolute value 1.*

There are infinitely many Pisot numbers. Indeed, if α is a Pisot number, then so are the powers α^n $(n = 1, 2, \dots)$, respectively. Moreover, the integers are trivially Pisot numbers, and $\frac{1+\sqrt{5}}{2}$ is one. It is easy to construct others using the theorem of Rouché: If f and g are two polynomials such that $|f(z) - g(z)| < |g(z)|$ for all z on a circle $C : |z| = R$, then f and g have the same number of zeroes (counting multiplicities) inside the circle $|z| < R$. (Since the inequality implies that $F := f/g$ satisfies $|F(z) - 1| < 1$ on C, i.e. the curve $F \circ C$ is contained in the open disk $|w - 1| < 1$, and hence its winding number $\int_C d\log F$ around 0 equals 0. But this winding number is the number of zeros minus the number of poles of f/g contained in $|z| < R$.)

Theorem 1.8. *Let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{Z}[x]$ such that*

$$1 + |a_{n-2}| + |a_{n-3}| + \cdots + |a_0| < |a_{n-1}|.$$

Then exactly one root α of f satisfies $\alpha| > 1$, and all other roots α' satisfy $|\alpha'| < 1$. In particular, α or $-\alpha$ is a Pisot number.

Proof. The inequality implies $|f(z) - a_{n-1}z^{n-1}| < |a_{n-1}z^{n-1}|$ on the circle $|z| = 1$. By Rouchés theorem f has thus $n - 1$ roots inside the unit disk $|z| < 1$, and since $|a_0| \geq 1$, it has then exactly one root α outside the unit circle. Since $\overline{\alpha}$ is also a root of f we have $\overline{\alpha} = \alpha$, hence α or $-\alpha$ is real and greater that 1. □

The smallest Pisot number has been determined by Siegel [Sieg]. It is the real root of

$$f_S = x^3 - x - 1$$

(see section 1.8).

For Salem numbers there is a construction due to Salem, also based on Rouché's theorem.

Theorem 1.9. *Let f be the minimal polynomial of a Pisot number of degree greater or equal to 3, let $\kappa = \pm 1$, and set $p_n = x^n f + \kappa f^*$. Then there is an n_0 such that, for any $n \geq n_0$, one root of p_n is a Salem number.*

Proof. We leave it as an exercise to show that there is some n_0 such that for all sufficiently small $\varepsilon > 0$ and all $n \geq n_0$ one has $|z^n p(z)/p^*(z)| > 1$, i,e. $|p_n(z) - z^n p(z)| < |z^n p(z)|$, on the circle $|z| = 1 + \varepsilon$. Hence p_n has $n + \deg p = \deg p_n - 1$ zeroes on $|z| \leq 1$, and exactly one, say α, outside the unit circle. Since $p_n^* = \pm p_n$, the set of zeroes of p_n is invariant under $z \mapsto 1/z$. Hence all zeros different from α and $1/\alpha$ must lie on the unit circle $|z| = 1$. □

It is not yet known whether there is a smallest Salem number. A proof (or disproof) of this fact would be an important contribution towards deciding the Lehmer conjecture. The smallest*known* Salem number can be obtained by Salem's construction:

$$z^7 f_S - f_S^* = x^8(x^3 - x - 1) - (-x^3 - x^2 + 1) = (x-1) f_L,$$

i.e. the unique root outside the unit circle of the polynomial f_L of Lehmer (see the end of last section), which has the so-far smallest known Mahler measure.

1.4 The height of an algebraic number

Before discussing further the Mahler measure and the known results in the direction of the Lehmer conjecture, we introduce its more number theoretic counter part, namely, the height of algebraic numbers. For an algebraic number α of degree n we define its "absolute" height by

$$H(\alpha) = \mu(f)^{1/n},$$

where f is the normalized minimal polynomial of α. The normalizing power $1/n$ is usually inserted to have a decomposition formula of the height function in local contributions which does depend on the field from which the valuations are taken. We shall explain this more precisely in a moment (see the proof of the next theorem).

If we set

$$f = a_n \prod_{j=1}^{n} (x - \alpha_j),$$

then

$$H(\alpha) = \Big[|a_d| \prod_{j=1}^{d} \max(1, |\alpha_j|) \Big]^{1/n}.$$

Note that this generalizes the height of a rational number defined in the first section. Indeed, if $\alpha = \frac{r}{s}$ with relative prime integers r, s, then $f = sx - r$, and hence $\mu(f) = |s|$ if $|r| \leq |s|$, and $\mu(f) = |r|$ otherwise.

As for the height of rational numbers one has a decomposition into local height contributions. We recall first of all the relevant facts about the valuations of an arbitrary number field K.

An equivalence class of valuations of K is called place of K or a prime of K. We always use P_K for the set of places of K, and we use P_K^∞ for the set of archimedean places of K, i.e. the set of equivalence classes of valuations which extend the usual absolute value on $\mathbb{Q}$ (up to equivalence).

The representatives $|\cdot|_v$ for the places v of K can be chosen in a unique way that one has

$$\prod_{v\in P(K)_\infty} |\alpha|_v = |\mathrm{N}_{K/\mathbb{Q}}(\alpha)|$$

and

$$\prod_{v\in P_K} |\alpha|_v = 1.$$

for all $\alpha \in K$. We always assume that $|\cdot|_v$ is normalized in this way.

One can describe the $|\cdot|_v$ explicitly as follows. To each prime ideal $\mathfrak{p}$ of K one can associate a valuation by

$$|\alpha|_\mathfrak{p} = \mathrm{N}_{K/\mathbb{Q}}(\mathfrak{p})^{-k},$$

where $\mathfrak{p}^k$ is the exact divisor of α. This valuation satisfies the stronger triangle inequality

$$|\alpha+\beta|_\mathfrak{p} \le \min(|\alpha|_\mathfrak{p}, |\beta|_\mathfrak{p})$$

with equality if $|\alpha|_\mathfrak{p}$ and $|\beta|_\mathfrak{p}$ are different.

Let $\sigma_j : K \mapsto \mathbb{R}$ $(1 \le j \le r)$ be the real embeddings of K, and let $\sigma_j, \overline{\sigma_j} : K \mapsto \mathbb{C}$ $(r < j \le r+s+1)$ be the pairs of complex embeddings of K. Then, for each j we have the valuation

$$|\alpha|_j = |\sigma_j(\alpha)|^{e_j},$$

where the bars on the right indicate the usual absolute value in $\mathbb{R}$ or $\mathbb{C}$, and where $e_j = 1$ if σ_j is real, and $e_j = 2$ if σ_j is complex.

It is a known fact that for any place v of K the valuation $|\cdots|_v$ equals $|\cdot|_\mathfrak{p}$ for some $\mathfrak{p}$ if it is finite (i.e. non-archimedean), and that it equals $|\cdot|_j$ for some j otherwise.

We shall also need the following two facts. Let L be a finite extension of K. Then the compatibility formula holds true, i.e.

$$|\alpha|_v^{[L:K]} = \prod_{\substack{w\in P_L\\ w|v}} |\alpha|_w$$

for all α in the ground field K. Here $w|v$ means that $|\cdot|_w$ is an extension of $|\cdot|_v$ up to equivalence. Example: For $5 = (1+2i)(1-2i)$ and $K = \mathbb{Q}(i)$ one finds

$$|5|_{(5)}^2 = |5|_{(1+2i)}|5|_{(1-2i)},$$

where on the left we have the 5-adic valuation on $\mathbb{Q}$, and on the right the corresponding valuations on $\mathbb{Q}(i)$.

If L is galois over K then, for any place v of K the Galois group $\mathrm{Gal}(L/K)$ acts transitively on the places w of L dividing v.

We are now in the position of proving the following decomposition formula for the absolute height.

Theorem 1.10. *Let K be a number field and $\alpha \in K$. Then one has*

$$H(\alpha) = \prod_{v \in P_K} \max(1, |\alpha|_v)^{1/[K:\mathbb{Q}]}.$$

Proof. Note first of all that the value of the right hand side does not depend on the field K. This is an immediate consequence of the compatibility formula and the fact that $|\alpha|_v < 1$ if and only if $|\alpha|_w < 1$ for all $w|v$ in any extension L of K.

The formula is trivial in the case that $K = \mathbb{Q}(\alpha)$ and α is integral. Indeed, in this case $|\alpha|_v \leq 1$ for all finite places v, and hence the $[K : \mathbb{Q}]$th power over the local contributions equals

$$\prod_{j=1}^{r} \max(1, |\sigma_j(\alpha)|) \prod_{j=r+1}^{s} |\max(1, \sigma_j(\alpha)|^2),$$

with σ_j having the same meaning as before. But this is exactly $\mu(f)$.

In the general case one can proceed as with the case of an integral α to prove

$$H(\alpha) = |a_n|^{1/[\mathbb{Q}(\alpha):\mathbb{Q}]} \prod_{v \in P_K^\infty} \max(1, |\alpha|_v)^{1/[K:\mathbb{Q}]},$$

where $|a_n|$ is the leading term of the normalized minimal polynomial of f. Hence it remains to relate $|a_n|$, the leading term of the normalized minimal polynomial of f to the factors associated to the finite primes of K.

For this one uses Gauss's lemma [Heck], p. 105:

$$\mathrm{cont}_v(g_1 g_2) = \mathrm{cont}_v(g_1)\,\mathrm{cont}_v(g_2).$$

for all $g_1, g_2 \in K[X]$, and all finite places v of the number field K. Here

$$\mathrm{cont}_v(a_m x^m + \cdots + a_0) = \max_j |a_j|_v.$$

By enlarging K if necessary, we may assume that K is Galois, say with Galois group G, and contains all roots of f. We can write f in the form

$$f = a_n \prod_{\sigma \in G} \big(x - \sigma(\alpha)\big)^{1/[K:\mathbb{Q}(\alpha)]}.$$

Let p be a rational prime number. We then have

$$\begin{aligned}
1 = \mathrm{cont}_p(f)^{[K:\mathbb{Q}(\alpha)]} &= |a_n|_p^{[K:\mathbb{Q}(\alpha)]}\,\mathrm{cont}_p\Big(\prod_{\sigma\in G}(x-\sigma(\alpha))\Big)\\
&= |a_n|_p^{[K:\mathbb{Q}(\alpha)]}\prod_{\sigma\in G}\prod_{v|p}\mathrm{cont}_v(x-\sigma(\alpha))^{1/[K:\mathbb{Q}]}\\
&= |a_n|_p^{[K:\mathbb{Q}(\alpha)]}\prod_{v|p}\prod_{\sigma\in G}\max(1,|\sigma(\alpha)|_v)^{1/[K:\mathbb{Q}]}\\
&= |a_n|_p^{[K:\mathbb{Q}(\alpha)]}\prod_{v|p}\max(1,|\alpha|_v)
\end{aligned}$$

Here the second identity follows from Gauss's lemma, the third one from the compatibility relation and Gauss lemma, and the last since G acts transitively on the places of K dividing p. Thus we find

$$|a_n| = \prod_{p\text{ finite}}\frac{1}{|a_n|_p} = \prod_{p\text{ finite}}\prod_{v|p}\max(1,|\alpha|_{\mathfrak{p}})^{1/[K:\mathbb{Q}(\alpha)]},$$

which implies the asserted formula. □

Sometimes one defines for a number field K the relative height function H_K on K by by

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

In particular one has

$$H_{\mathbb{Q}(\alpha)}(\alpha) = \mu(f)$$

, where f is the normalized minimal polynomial of α.

Since the Mahler measure and the height of algebraic numbers represent essentially the same notion, one can easily deduce several properties of the height from the Mahler measure.

For instance, for any degree d and any bound B there is only a finite number of $\alpha\in\overline{\mathbb{Q}}$ of degree less or equal to d with $H(\alpha)\leq B$. Moreover

$$H(\alpha) = H(1/\alpha)$$

for any $\alpha\neq 0$ (since $\pm f^*$ is the normalized minimal polynomial of $1/\alpha$ if f is the normalized minimal polynomial of α. Kronecker's theorem states states that, for any algebraic α one has $H(\alpha)=1$ if and only if α is a root of unity.

Finally, Lehmer's conjecture is equivalent to the fact that for some constant $C>1$ one has

$$H(\alpha)\geq C^{\deg\alpha}$$

for all algebraic α. Note the degree $\deg\alpha$ of α in this formula. Without this degree such an inequality cannot be true. Counter example:

$$H(2^{1/n}) = 2^{1/n} \mapsto 1.$$

1.5 Two easy Lehmer type theorems

In this section we prove a theorem of Schinzel from 1973 concerning an absolute lower bound for the height of totally real algebraic numbers, and a more recent one one of Zhang from 1992 about numbers simultaneously "close to 0 and 1". Both theorems admit surprisingly simple proofs ([HoSk] and [Zagi]) which are quite similar though they were found independently [1] . In this section we give the the proofs without any additional comment. A reinterpretation and two possible generalizations will follow in the two next sections.

Theorem 1.11. *(Schinzel [Schi]) Let $\alpha \neq 0, \pm 1$ be a totally real algebraic number. Then*

$$H(\alpha) \geq \sqrt{\frac{1+\sqrt{5}}{2}} = 1.2720\ldots,$$

with equality if and only if α equals one of the four numbers $\pm\frac{1\pm\sqrt{5}}{2}$.

Note that a theorem like this cannot be true in general, i.e. there is no absolute lower bound for the absolute height of all but a finite number of algebraic numbers. Indeed, $x^p - a$ is irreducible for all square-free positive integers a, and all rational primes p. Thus

$$H(\sqrt[p]{a}) = a^{1/p} \to 1.$$

Proof. (cf. [HoSk]) If, for x real, we set $\gamma(x) = |x|^{1/2}|x - 1/x|^{1/2\sqrt{5}}$, then we have

$$\max(1,|x|) \geq \sqrt{\frac{1+\sqrt{5}}{2}}\,\gamma(x),$$

with equality if and only if $x = \pm\frac{1\pm\sqrt{5}}{2}$. Indeed, since $|x|\gamma(1/x) = \gamma(x)$, since the same invariance property holds for the function $\max(1,|x|)$, and since both sides of the desired inequality are invariant under $x \mapsto -x$, it suffices

[1] The second proof differs slightly from the original version given in [Zagi]. When I prepared this manuscript I noticed that Zagier's proof could be presented in a form which makes it look much more similar to the proof of Schinzel's theorem in [HoSk].

to prove it for $0 \leq x \leq 1$. But in this interval maximum of $\gamma(x)$ occurs for $x = \frac{-1+\sqrt{5}}{2}$ with maximum value $\sqrt{\frac{-1+\sqrt{5}}{2}}$.

On the other hand

$$|a| \prod_j \gamma(\alpha_j) = |a|^{1/2-1/2\sqrt{5}} |f(0)|^{1/2-1/2\sqrt{5}} |f(1)f(-1)|^{1/2\sqrt{5}} \geq 1$$

where $f(x) = a \prod(x - \alpha_j)$ is the minimal polynomial of α. The result is now obvious. □

Theorem 1.12. *(Zhang [Zhan]) For all algebraic numbers $\alpha \neq 0, 1, \frac{1\pm\sqrt{-3}}{2}$, one has*

$$H(\alpha)H(1-\alpha) \geq \sqrt{\frac{1+\sqrt{5}}{2}}$$

with equality if and only if α or $1-\alpha$ is a primitive 10th root of unity.

Proof. (Cf. [Zagi]) Here, for complex z, we set

$$\gamma(z) = |z|^{1/2} |1-z|^{1/2} \Big(\frac{|z^2 - z + 1|}{|z^2 - z|} \Big)^{1/2\sqrt{5}}.$$

It is straight-forward, though cumbersome, to prove

$$\max(1, |x|) \max(1, |1-x|) \geq \sqrt{\frac{1+\sqrt{5}}{2}} \, \gamma(x)$$

for all complex arguments, with equality if and only if x or $1-x$ equals $\mathrm{e}^{\pm\pi i/5}$ or $\mathrm{e}^{\pm 3\pi i/5}$. With the same notations as in the theorem before we find

$$|a| \prod_j \gamma(\alpha_j) = |f(0)f(1)|^{1/2-1/2\sqrt{5}} |f(\frac{1+\sqrt{-3}}{2}) f(\frac{1-\sqrt{-3}}{2})|^{1/2\sqrt{5}} \geq 1,$$

which again implies the result. □

1.6 Numbers with conjugates outside a given set

The proofs of the two theorems of the preceding section have obviously very much in common. The both use a function $\gamma(z)$ with remarkable symmetry to bound $\max(1, |z|)$ to below. We formalize the construction of this bounding function.

Fix an arbitrary polynomial $p \neq 0$ with integral coefficients and a real number $s > 0$, and set

$$\gamma(z) := |z|^{1/2} |p(z)p(1/z)|^s.$$

Then

$$|z|\gamma(1/z) = \gamma(z).$$

Note that we also have

$$|z| \max(1, |1/z|) = \max(1, zx|).$$

Moreover, $\gamma(z)$ and $\max(1, |z|)$ are both invariant under $z \mapsto \overline{z}$. Thus if

$$\max(1, |z|) \geq \gamma(z)$$

for z in some subset E of $\mathbb{C}$, then we can assume without loss of generality that E is invariant under $z \mapsto 1/z$ and $z \mapsto \overline{z}$. By continuity we can furthermore assume that E is closed. The invariance of $z \mapsto 1/z$ implies that, for proving the desired estimate, we only have to look at arguments z in the intersection of E with the unit disk $|z| \leq 1$. Using that E is stable under complex conjugation it even suffices to look at the intersection of E with the unit circle. More precisely we have the following lemma.

Lemma 1.1. *Let E be a closed subset of $\mathbb{C}$ invariant under complex conjugation and under $z \mapsto 1/z$. Let $p \in \mathbb{C}[x]$, and suppose that*

$$\sup_{z \in E,\ |z|=1} |p(z)| < 1.$$

Then, for all sufficiently small $s > 0$ there exists a constant $C > 1$ such that

$$\max(1, |z|) \geq C\,|z|^{1/2} |p(z)p(1/z)|^s.$$

for all $z \in E$.

Proof. Denote by $\mathbb{D}$ the closed unit disk $|z| \leq 1$. We claim that there is a non-negative integer l such that

$$a := \sup_{z \in E \cap \mathbb{D}} |z^l p^*(z) p(z)| < 1.$$

Indeed, otherwise there would be a sequence z_k in $E \cap \mathbb{D}$ and of non-negative integers n_k such that $|z_k^{n_k} p^*(z_k) p(z_k)| \geq 1$. Since $E \cap D$ is compact, we may assume that z_k converges towards an $w \in E \cap D$. For this w we have by continuity $|p(w)p^*(w)| \geq 1$. If we had $|w| = 1$ then $|p(w)| < 1$, hence

$|p^*(w)| > 1$. But the latter is impossible since $|p^*(w)| = |p(\overline{w})|$ for $|w| = 1$, and since $\overline{w} \in E$ by the invariance of E under complex conjugation. But then $|w| < 1$, and hence $|p(z_k)p^*(z_k)| \geq |z_k|^{-n_k} \to \infty$, which is absurd.

Thus, for any $s \leq 1/2(l + \deg p)$ there is a $C > 1$ such that the desired estimate holds for all $z \in E \cap \mathbb{D}$. But this holds then true for all $z \in E$ by the transformation formulas of both sides of the inequality under $z \mapsto 1/z$. □

The lemma explains how to find a function $\gamma(x)$ as used for instance in the proof of Schinzel's theorem: There $E = \mathbb{R}$, thus any integral polynomial p with $|p(x)| < 1$ for $x \in \{\pm 1\}$, the intersection of $\mathbb{R}$ with the unit circle, would lead to a lower bound $C > 1$ for totally real numbers. The polynomial $p(x) = x^2 - 1$ is certainly the simplest solution, and it it exactly the one which one finds in our proof.

Now suppose finally that we have an integral $p \neq 0$ satisfying the hypothesis of the lemma. Then this yields immediately an absolute lower bound for the heights of algebraic numbers all of whose conjugates lie outside E.

Lemma 1.2. *Let E, p and $C > 1$ as in the preceding lemma. Suppose furthermore that $p \in \mathbb{Z}[x]$ and $p \neq 0$. Then*

$$H(\alpha) \geq C$$

for all α such that α and all its conjugates are contained in E and different from 0 and the roots of p and p^*..

Since $H(\alpha) = 1$ for roots of unity α, we see that the existence of a p satisfying the hypothesis of the lemma implies that, for each integer $n \geq 1$, either $\mathbb{C} \setminus E$ contains at least one primitive root of unity, or else p has all nth roots of unity as roots. In particular, $\mathbb{C} \setminus E$ must intersect the unit circle non-trivially.

Proof. By the preceding lemma we have the following estimate:

$$H(\alpha)^n C^{-n} \geq |a| \prod_{j=1}^{n} \gamma(\alpha_j) = |a|^{1/2-2ls} |f(0)|^{1/2-ls} a^{ls} \prod_{j=1}^{n} |p(\alpha_j)p^*(\alpha_j)|^s,$$

where l is the degree of p. But $|f(0)|$ and the factor after it are positive powers of positive integers. Hence, for $s < 1/4l$, we obtain

$$H(\alpha) \geq C.$$

□

It is certainly natural to start now with a set E, and to ask when we can find an integral $p \neq 0$ satisfying the assumptions of the first lemma. The answer will be found using the theory of transfinite diameters whose basics we shall develop in the next section. We shall find as answer:

Lemma 1.3. *Let E be a closed subset of $\mathbb{C}$ not containing the whole unit circle and stable under complex conjugation. Then there is polynomial $p \in \mathbb{Z}[x]$, $p \neq 0$, such that*

$$\sup_{z\in E,\ |z|=1} |p(z)| < 1.$$

Note that the condition that E does not contain the unit circle, is also necessary. Otherwise we would be able to prove that $H(\alpha)$ is absolutely bounded below by a constant greater than 1 for all α which are not roots of unity or 0, which is certainly false (see the counter example $H(\sqrt[n]{2})$ at the end of section 1.4. Postponing the proof of the preceding lemma to the next section we can summarise by saying:

Theorem 1.13. *(Langevin [Lvin]) Let G be an open region in $\mathbb{C}$ which intersect the unit circle $|z| = 1$. Then there exists a constant $C(G) > 1$ such that*

$$H(\alpha) \geq C(G)$$

for any $\alpha \in \overline{\mathbb{Q}}$ which has no conjugates in G, which is not a root of unity and different from 0.

Proof. This is a consequence of the preceding lemmas. For applying the lemma 1.1 we actually need that G, or equivalently $E := \mathbb{C} \setminus G$, is invariant under complex multiplication and $z \mapsto 1/z$. However, this can be assumed without loss of generality. If z_0 is on the unit circle and in G, then there is an open neighbourhood of z_0 and contained in G which is stable under $z \mapsto 1/\overline{z}$ (exercise). Clearly, we may replace G by this neighbourhood. Secondly, we may then replace G by the union of G and $G^* := \{\overline{z} : z \in G\}$, since all conjugates of α are outside G if and only if they are outside G^*. The resulting G has then the necessary invariance properties.

By the last lemma we find, for E, a p satisfying the hypothesis of Lemma 1.1. Thus Langevin's theorem is true for all $\alpha \neq 0$ having no conjugates in G apart from possibly those roots β of p or p^* which are not roots of unity. (Note in passing that any lth root of unity has a conjugate in G if $l \gg 0$, and that any root of unity for which this does not hold true must be a root of p). However, the heights of the β are strictly larger than 1. Hence by choosing a $C(G) > 1$ which is smaller than these finitely many heights and smaller than C, we obtain the desired theorem. □

There is a somewhat remarkable consequence of Langevin's theorem, which suggests that it is easier to believe Lehmer's conjecture than the contrary.

Corollary 1.13.1. *If there were a sequence f_n of polynomials in $\mathbb{Z}[x]$ with $\mu(f_n) > 1$ but $\lim \mu(f_n) = 1$ (i.e. if the Lehmer conjecture where false), then any point on the unit circle $|z| = 1$ would be an accumulation point of the roots of the f_n.*

1.7 Transfinite diameters

In this section we shall prove the main lemma 1.3, which we needed for the proof Langevin's theorem. For this we shall need the theory of transfinite diameters invented by Fekete and Tonelli [FeTo].

For a compact subset E of $\mathbb{C}$ we set

$$\rho_n(E) := \inf\{|p|_E : p \in \mathbb{C}[x]_n,\ p = x^n + \dots\}.$$

Here we use

$$|p|_E = \sup_{z \in E} |p(z)|.$$

One can show that the number $\rho_n(E)$ is even attained by a unitary polynomial $C_n \in \mathbb{C}[x]_n$, and that, even more, C_n is unique if E has more than n points [FeTo]. This C_n is called the nth Chebyshev polynomial of E.

Note that, in the definition of $\rho_n(E)$, we can restrict to real polynomials if E is stable under complex conjugation. Namely, in this case we have

$$|\frac{1}{2}(p + \overline{p})|_E \leq \frac{1}{2}(|p|_E + |\overline{p}|_E) = |p|_E$$

for any polynomial p, where $\overline{p}$ is obtained from p by taking the complex conjugates of the coefficients of p. In particular, C_n is real for such E.

By a similar argument we see that, in the case that E is the unit circle $\mathbb{S} : |z| = 1$, we can restrict to those unitary polynomials in $\mathbb{C}[x]_n$ which are invariant under $x \mapsto \zeta x$ for all nth roots of unity. Indeed,

$$\tilde{p}(x) = \frac{1}{n} \sum_{\zeta^n = 1} p(\zeta x)$$

is unitary of degree n and satisfies $|\tilde{p}|_{\mathbb{S}} \leq |p|_{\mathbb{S}}$. On the other hand, the only unitary polynomials in $\mathbb{C}[x]_n$ which are invariant under all $x \mapsto \zeta x$, are the

$x^n + c$, where c is a constant (exercise). Clearly x^n is thus the nth Chebyshev polynomial of the unit circle, and consequently

$$\rho_n(\mathbb{S}) = |x^n|_{\mathbb{S}} = 1.$$

For an integer $n \geq 0$ let $T_n(x)$ be the polynomial defined by

$$\cos(nt) = 2^{n-1} T_n(\cos t).$$

Thus T_n is a unitary polynomial of degree n. One has $T_1 = x$, $T_2 = x^2 - 1$, $T_3 = x^3 - \frac{3}{4}x$. The polynomial $2^{n-1}T_n$ is the polynomial which is usually called the nth Chebishev polynomial without making any reference to transfinite diameters. In fact, T_n is the nth Chebishev polynomial of the interval $I = [-1, +1]$ in the sense defined before.

Indeed, as is obvious from the definition, $|T_n(x)|_I = 1/2^{n-1}$. Furthermore $T_n(x)$ attains $n + 1$ times the critical values $\pm 2^{1-n}$ in the interval $[-1, +1]$, with alternating signs from left to right. Hence, if p were a unitary polynomial of degree n with $|p|_I < 2^{1-n}$, then $f - p$ would be a polynomial different from 0, of degree $\leq (n - 1)$ and whose values changes $n + 1$ times the sign in I; thus it would have n zeroes, a contradiction.

The transfinite diameter (or Chebichev constant) of a compact subset E is defined as

$$\rho = \lim_n \rho_n(E)^{1/n}.$$

For the unit circle and closed intervals $[a, b]$ on the real axis we have by the preceding discussion

Theorem 1.14. *One has $\rho(\mathbb{S}) = 1$ and $\rho([a, b]) = \frac{b-a}{4}$ for all real $a \leq b$.*

Proof. The second formula follows from $\rho([-1, +1]) = \frac{1}{2}$ on using the general formulas $\rho(t + E) = \rho(E)$ and $\rho(\lambda E) = \lambda\rho(E)$ (which, in turn are an obvious consequence of $|f(x)|_{t+E} = |f(x + t)|_E$ and $|f(x)|_{\lambda E} = |\lambda||\lambda^{-1} f(\lambda x)|_E$). □

Theorem 1.15. *The limit $\rho(E)$ exists and is finite.*

Proof. Since E is compact it is contained in a disk $|z| \leq R$ for some R. Hence $\rho_n(E)^{1/n} \leq |x^n|_E^{1/n} = R$. In particular,

$$\alpha := \liminf \rho_n(E)^{1/n}, \qquad \beta := \limsup \rho_n(E)^{1/n}$$

are both finite. Let $\varepsilon > 0$, and let p a unitary polynomial, say of degree l, such that $|p|_E^{1/l} < \alpha + \varepsilon$. Then there exists a constant C such that

$$|z^r p^k|_E \leq C(\alpha + \varepsilon)^n \quad (n = kl + r,\ 0 \leq r < n)$$

for all n. Hence $\rho_n(E)^{1/n} \leq C^{1/n}(\alpha + \varepsilon) \to \alpha + \varepsilon$, and hence $\beta = \alpha$. □

We shall need

Lemma 1.4. *Let A be an arc of length $t \leq 2\pi$ on the unit circle. Then*

$$\rho(A) \leq \sin(t/4).$$

In fact one can show that one actually has equality [FeTo]

Proof. By applying a suitable rotation we may suppose that A is stable under complex conjugation, contains 1 and has end points $\mathrm{e}^{it/2}$ and $\mathrm{e}^{-it/2}$. Consider the map

$$R : A \to I := [\cos(t/2), 1], \quad z \mapsto \frac{1}{2}(z + \frac{1}{z}),$$

which is 2 to 1. If p is a unitary polynomial of degree n, then

$$|p|_I = |p \circ R|_A = 2^{-n}\big|(2x)^n p\big(\frac{1}{2}(x + \frac{1}{x})\big)\big|_A \geq 2^{-n}\rho_{2n}(A),$$

and hence

$$\frac{1 - \cos(t/2)}{4} = \rho(I) \geq \rho(A)^2,$$

i.e. $\sin^2(t/4) \geq \rho(A)^2$. □

Theorem 1.16. *(Kakeya) Let E be a compact subset such that $|p|_E < 1$ for some unitary polynomial $p \in \mathbb{R}[x]$. Then there exists a unitary polynomial $q \in \mathbb{Z}[x]$ with $|q|_E < 1$.*

Proof. Let p be real, unitary, say of degree n with $|p|_E < 1$. Clearly $n \geq 1$. For positive integral m write $m = qn + r$ with integral q and $0 \leq r < n$, and set

$$p_m(x) = x^r p(x)^q.$$

Then

$$|p_m| \leq b\, a^m \qquad (b = \max_{0 \leq r < n} |z^r|_E,\ a = |p|_E^{1/n}.)$$

Fix a positive integer s. For each r we can find scalars $|\lambda_j| < 1$ such that

$$L_r := p_r + \lambda_1 p_{r-1} + \cdots + \lambda_{r-s} p_s = G_r + H_r$$

with a unitary $G_r \in \mathbb{Z}[x]$ and an H of degree strictly smaller than s and whose coefficients have absolute value less than 1. One has

$$|L_r|_E \leq b(a^r + \cdots + a^s) \leq b\frac{a^s}{1 - a}.$$

Thus if s is sufficiently big, then $|L_r|_E < 1/3$ for all r. But by construction the sequence $|H_r|_E$ is bounded. Hence for some $r_1 > r_2$ we have $|H_{r_1} - H_{r_2}|_E < 1/3$. But then

$$|G_{r_1} - G_{r_2}|_E = |L_{r_1} - H_{r_1} - (L_{r_2} - H_{r_2})|_E \leq |L_{r_1}|_E + |L_{r_1}|_E + |H_{r_1} - H_{r_2}|_E < 1.$$

□

1.8 Heights of non-reciprocal numbers

We call a polynomial reciprocal if its set of roots is invariant under $z \mapsto 1/z$, and and we call an algebraic number $\alpha \neq 0$ reciprocal if the set of all conjugate numbers is invariant under $z \mapsto 1/z$. Clearly, a polynomial is reciprocal if and only if if and only if $f^* = af$ for some number a.

Theorem 1.17. *(Smyth) Let $f \in \mathbb{Z}[x]$, and assume that f is not reciprocal and $f(0) \neq 0$. Then*

$$\mu(f) \geq \theta = 1.3247\ldots,$$

where θ is the real solution of $\theta^3 - \theta - 1 = 0$.

Corollary 1.17.1. *If $f \in \mathbb{Z}[x]$ is irreducible and of odd degree, then*

$$\mu(f) \geq \theta.$$

Proof. Assume that $f = af^*$ for same integer a. Then the set of roots of f is invariant under the involution $\alpha \mapsto 1/\alpha$. Hence, if the degree of f were odd, then at least one root satisfies $\alpha = 1/\alpha$, i.e. $\alpha = \pm 1$. Hence any irreducible polynomial of odd degree is either equal to a multiple of $x+1$ or $x-1$, or else is not reciprocal. Hence Smyth theorem applies to all irreducible polynomials of odd degree. □

We may restate Smyth theorem and its corollary by saying: If α is an algebraic number of degree n such that n is odd degree or such that α is not reciprocal, then

$$H(\alpha) \geq \sqrt[n]{\theta}.$$

Corollary 1.17.2. *(Siegel) θ is the smallest Pisot number.*

Proof. The set of roots of a minimal polynomial f of a Pisot number can only be invariant under $\alpha \mapsto 1/\alpha$ if the degree of f is two. Thus Smyth theorem applies to f unless f is of degree 2. But in the latter case $\mu(f) \geq \frac{1+\sqrt{5}}{2}$ as we already saw before (see section 1.2. □

Corollary 1.17.3. *(Cassels) Assume that $f(x) = \prod_{j=1}^{n}(x - \alpha_j) \in \mathbb{Z}[x]$ satisfies $|\alpha_j| < 1 + \frac{\log\theta}{n}$ $(1 \leq j \leq n)$. Then $f = \pm f^*$.*

Proof. We remark that the original theorem of Cassels was stated with $\log\theta = 0.28\ldots$ replaced by 0.1.

For the proof we simply note that by assumption

$$\mu(f) < \left(1 + \frac{\log\theta}{n}\right)^n \leq \mathrm{e}^{\log\theta} = \theta.$$

Thus Smyth theorem cannot apply to f. □

1.9 Proof of Smyth's theorem

We follow in this section essentially the original proof in [Smy1]. For a complex number α set

$$B_\alpha(z) = \frac{z-\alpha}{1-\overline{\alpha}z}.$$

If α is inside the unit disk then B_α is holomorphic in an open neighborhood of the unit disk and satisfies $|B_\alpha(z)| = 1$ for $|z| = 1$. Let now $\alpha_1, \ldots, \alpha_r$ be complex numbers inside the unit disk, and let

$$B(z) = \prod_{j=1}^{r} B_{\alpha_j}(z) = c_0 + c_1 z + c_2 z^2 + \cdots .$$

We shall call B the Blaschke function associated to the family of the a_j.

Lemma 1.5. *One has* $1 = |c_0|^2 + |c_1|^2 + |c_2|^2 + \cdots$,

Proof. This follows from

$$1 = \frac{1}{2\pi}\int_0^{2\pi} |B(\mathrm{e}^{it})|^2 \, dt = \frac{1}{2\pi}\sum_{k,l} c_k c_l \int_0^{2\pi} \mathrm{e}^{i(k-l)} \, dt.$$

□

Assume now that f is a real polynomial without zeroes on the unit circle and such that $f(0) \neq 0$. Let B and $\hat{B}$ be the Blaschke functions associated to the zeroes of f and f^* inside the unit circle, respectively (with repeated multiple roots). Then B/f has no zeroes, and its poles are the roots of f outside the unit circle and the $1/\overline{\alpha}$, where α runs through the roots of f inside the unit circle; and the same holds true for $\hat{B}/f^*$ since the set of roots of f is invariant under $z \mapsto \overline{z}$. In fact, one easily checks

$$\frac{B}{f} = \frac{\hat{B}}{f^*}.$$

Let c_k, d_k and a_k denote the Taylor coefficients of B, $\hat{B}$ and f/f^* at $z = 0$, respectively. Assume that the constant and leading term of f are equal to ± 1. Then $c := |c_0| = |d_0| = 1/\mu(f)$ and $a_0 = \pm 1$. If, furthermore, f/f^* is not constant, then there exists a smallest $k \geq 1$ such that $a_k \neq 0$, and consequently,

$$c_k - a_0 d_k = a_k d_0.$$

From this (and $|a_0| = 1$) we see that $|d_k| \geq |a_k d_0|/2 = |a_k|c/2$ or $|c_k| \geq |a_k|c/2$. Hence from the preceding lemma

$$1 \geq c^2 + \frac{|a_k|^2}{4}c^2,$$

and thus

$$\mu(f) \geq \Big(1 + \frac{|a_k|^2}{4}\Big)^{\frac{1}{2}}.$$

Assume now that f is integral, irreducible and not reciprocal and $f(0) \neq 0$. Then it has no roots on the unit circle (since such a root would be a root of f^* too), and f/f^* is not constant. For the proof of Smyth theorem we may moreover assume that the leading term and constant term of f is 1 (since otherwise $\mu(f) \geq 2$). Thus f satisfies the hypothesis used in the last paragraph, and accordingly the last estimate for $\mu(f)$ holds true. However, here f/f^* has integral Taylor coefficients, in particular $|a_k| \geq 1$. Hence

$$\mu(f) \geq \sqrt{\frac{5}{4}} = 1.118\cdots.$$

This is already a weak version of Smyth's theorem. His sharp bound is obtained, essentially by the same method, However, by a more subtle investigation of the coefficients of the Blaschke function than in our lemma above.

Lemma 1.6. *Let $n \geq 1$.*

1. *For all real $x_0, \ldots, x_n$ one has*

$$\begin{aligned}(c_0x_0)^2 + (c_0x_1 + c_1x_n)^2 + \cdots + (c_0x_n + \cdots + c_nx_0)^2 \\ \leq x_0^2 + x_1^2 + \cdots + x_n^2. \quad (1.1)\end{aligned}$$

2. *Set*

$$A = \begin{pmatrix} c_n & c_{n-1} & c_{n-2} & \cdots & c_0 \\ c_{n-1} & c_{n-2} & \cdots & c_0 & 0 \\ c_{n-2} & \cdots & c_0 & 0 & 0 \\ \vdots & & & & \\ c_0 & & & & \end{pmatrix}.$$

Then $1 + A$ and $1 - A$ are symmetric, positive definite matrices.

3. In particular, one has

$$1 \geq c_0^2 + |c_n|, \tag{1.2}$$

$$-(1 - c_0^2 - \frac{c_n^2}{1+c_0}) \leq c_{2n} \leq 1 - c_0^2 - \frac{c_n^2}{1-c_0}. \tag{1.3}$$

and the same inequalities with c_k replaced by d_k.

Proof. In fact, the above inequalities hold true for the Taylor coefficients at 0 of any function which is holomorphic in an open neighborhood of the unit disk $|z| \leq 1$, satisfies $|f(z)| \leq 1$ for $|z| = 1$ and has real Taylor coefficients c_j. Indeed, setting $p(z) = x_0 + x_1 z + \cdots + x_n z^n$, we have

$$\sum_{j=0}^{n} (c_0 x_j + \cdots + c_j x_0)^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(z)p(z)|^2 \, dt \qquad (z = \mathrm{e}^{it})$$
$$\leq \frac{1}{2\pi} \int_0^{2\pi} |p(z)|^2 \, dt = \sum_{j=0}^{n} x_j^2.$$

The second assertion is obtained using the Cauchy-Schwartz inequality and the first one:

$$\pm x^t A x \leq |x| \, |Ax| \leq |x|^2.$$

Let $\varepsilon = \pm 1$. Since $1 + \varepsilon A \geq 0$ we obtain in particular

$$\det \begin{pmatrix} 1+\varepsilon c_n & \varepsilon c_0 \\ \varepsilon c_0 & 1 \end{pmatrix}, \det \begin{pmatrix} 1+\varepsilon c_{2n} & \varepsilon c_n & \varepsilon c_0 \\ \varepsilon c_n & 1+\varepsilon c_0 & 0 \\ \varepsilon c_0 & 0 & 1 \end{pmatrix} \geq 0,$$

which implies the last two inequalities. □

We saw above that $\max(|c_n|, |d_n|) \geq |c|/2$. Together with the third inequality this implies already $1 - c^2 \geq c/2$, $\mu(f)^2 - \mu(f)/2 - 1 \geq 0$, and hence

$$\mu(f) \geq \frac{1+\sqrt{17}}{4} = 1.280\ldots.$$

We can assume without loss of generality that $\mu(f) \leq \frac{4}{3}$. We set

$$f(0)\frac{f}{f^*} = 1 + a_k z^k + a_l z^l + O(z^{l+1}),$$

where $k < l$ and $a_k, a_l \neq 0$. By multiplying B by $\pm f(0)$ and $\widehat{B}$ by ± 1 we can then assume that $B(0), \widehat{B}(0) > 0$ and that

$$B = (1 + a_k z^k + a_l z^l + \cdots)\widehat{B}.$$

In particular, we have

$$c := c_0 = d_0 = \mu(f)^{-1},$$

and furthermore

$$c_j = d_j \qquad (0 \leq j < k) \tag{1.4}$$

$$c_k = d_k + c_0 a_k \tag{1.5}$$

$$c_{k+1} = d_{k+1} + d_1 a_k \tag{1.6}$$

$$c_{l-1} = d_{l-1} + d_{l-k-1} a_k \tag{1.7}$$

$$c_l = d_l + d_{l-k} a_k + c_0 a_l \tag{1.8}$$

As a first consequence we note

$$|a_k| = 1 \tag{1.9}$$

$$|a_l| = 1 \tag{1.10}$$

$$|c_k| + |d_k| = c \tag{1.11}$$

Indeed, if $|a_k| \geq 2$, then, by (1.5), we would have $\max(|c_k|, |d_k|) \geq c$. Hence, by the lemma $1 - c^2 \geq c$, which contradicts $c \geq \frac{4}{3}$.

Similarly, if $|a_l| \geq 2$, then, by (1.8), $\max(|c_l|, |d_l|, |d_{l-k}|) \geq \frac{2}{3}c$, and hence, by the lemma, $1 - c^2 \geq \frac{2}{3}c$. Again this contradicts $c \geq \frac{4}{3}$.

Finally, by (1.5) $|c_k| + |d_k| \geq |c_k - d_k| \geq c$. If the inequality were strict, then $c = |c_k| - |d_k|$ or $c = |d_k| - |c_k|$, in any case, $\max(|c_k|, |d_k|) \geq c$, which is impossible as we have already seen.

Case $2k \leq l$: We can assume that $2k < l$. Otherwise we interchange f and f^*. Namely, using

$$(1 + a_k x^k + a_{2k} x^{2k} + \cdots)^{-1} = 1 - a_k x^k + (a_k^2 - a_{2k}) x^{2k} + \cdots,$$

we see that then $a_k^2 - a_{2k} = 0$ (since otherwise this would be 2 by (1.9), (1.10)).

We now apply (1.3) to obtain

$$-(1 - c^2 - \frac{c_k^2}{1 + c_0}) \leq c_{2k} \leq 1 - c^2 - \frac{c_k^2}{1 - c_0}$$

$$-(1 - d^2 - \frac{d_k^2}{1 - d_0}) \leq -d_{2k} \leq 1 - d^2 - \frac{d_k^2}{1 + d_0}).$$

Adding both inequalities gives (on using also (1.5))

$$-2(1 - c^2) + \frac{d_k^2}{1 - d_0} + \frac{c_k^2}{1 + c_0} \leq c_{2k} - d_{2k} = d_k a_k \tag{1.12}$$

$$\leq 2(1 - c^2) - \frac{d_k^2}{1 + d_0} - \frac{c_k^2}{1 - c_0}. \tag{1.13}$$

Using (1.11) this gives

$$|d_k| \leq \max(H(|d_k|), H(|c_k|)),$$

where we use

$$H(x) := 2(1-c^2) - \Big(\frac{x^2}{1+c} + \frac{(c-x)^2}{1-c}\Big).$$

But $1-c^2 \geq |d_k| \geq c-|c_k| \geq c+c^2-1$ by (1.3),(1.7) and (1.2), respectively, and the same holds true with c_k and d_k interchanged. Hence if we set $I = [c^2+c-1, 1-c^2]$, then we find

$$c^2+c-1 \leq \max_{x\in I} H(x).$$

But $H(x)$ takes its maximum in $x = \frac{1+c}{2}$. Since $c \geq \frac{3}{4}$ we have $\frac{1+c}{2} \geq 1-c^2$ (since the latter is equivalent to $c \not\in]-1, \frac{1}{2}[$). Hence $H(x)$ is increasing on I, and thus

$$c^2+c-1 \leq H(1-c^2) = 2(1-c^2) - \frac{(1-c^2)^2}{1+c} - \frac{(c^2+c-1)^2}{1-c},$$

i.e. $-c^3-c^2+1 \geq 0$. This gives finally $\mu(f)^3-\mu(f)-1 \geq 0$, which means that $\mu(f)$ is to the right of the real root θ of $x^3-x-1=0$.
Case $l < 2k$: We may assume $a_k = \pm 1$ (otherwise interchange f and f^*). By (1.1) we have, for all $\beta, \gamma \in \mathbb{R}$,

$$\begin{gathered} c^2 + (c_{l-k}+\gamma c)^2 + (c_k+\gamma c_{2k-l}-c)^2 + (c_l+\gamma c_k - c_{l-k}+\beta c)^2, \\ c^2 + (-d_{l-k}-\gamma c)^2 + (-d_k-\gamma d_{2k-l}-c)^2 + (-d_l-\gamma d_k - d_{l-k}+\beta c)^2 \\ \leq 2+\gamma^2+\beta^2. \end{gathered}$$

We add these two inequalities, use $\frac{a^2+b^2}{2} \geq \left(\frac{a+b}{2}\right)^2$ and $c_j = d_j$ for $1 \leq j < k$, and set $x = c_{l-k} = d_{l-k}$ to obtain

$$c^2 + \big(x+\gamma c\big)^2 + \big(\frac{c_k-d_k}{2}-c\big)^2 + \big(\frac{c_l-d_l}{2} + \gamma\frac{c_k-d_k}{2} - x+\beta c\big)^2 \leq 2+\gamma^2+\beta^2.$$

By (1.5) and (1.8), using $a_k = +1$, this can be rewritten as

$$\frac{5}{4}c^2 + (x+\gamma c)^2 + \big(\frac{x+ca_l}{2} + \gamma\frac{c}{2} - x + \beta c\big)^2 \leq 2+\gamma^2+\beta^2.$$

Replacing x by $-a_l x$, β by $a_l\beta$ and γ by $-a_l\gamma$ we get

$$\frac{5}{4}c^2 + (x+\gamma c)^2 + \big(\frac{c+x-\gamma c}{2} + \beta c\big)^2 \leq 2+\gamma^2+\beta^2.$$

If we view the difference of the right hand side and the left hand side as quadratic polynomial in β, then the inequality states that its discriminant is ≤ 0, i.e. (using $1 - c^2 > 0$) that

$$\frac{5}{4}c^2 + (x + \gamma c)^2 + \frac{(c + x - \gamma c)^2}{4(1 - c^2)} \leq 2 + \gamma^2.$$

Again, viewing the difference of both sides as polynomial in γ, we obtain that its discriminant is ≤ 0. Thus (using that the coefficient of γ^2 is positive since $c < 4/(1 + \sqrt{17})$, as follows from $1 - c^2 \geq c/2$) we have

$$\frac{5}{4}c^2 + x^2 + \frac{(c - x)^2}{4(1 - c^2)} + \frac{(2xc - \frac{c(c+x)}{2(1-c^2)})^2}{4(1 - c^2 - \frac{c^2}{4(1-c^2)})} \leq 2$$

Now, again, since $c < 4/(1 + \sqrt{17})$, the left hand side minus 2 viewed as polynomial in x has positive leading term. Since it is ≤ 0 for at least one x it has a real root, hence non negative discriminant. By a straight forward calculation this yields $40c^4 - 93c^2 + 40 \geq 0$, or, in terms of $\mu(f)$, finally

$$\mu(f)^4 - \frac{93}{40}\mu(f) + 1 \geq 0.$$

This implies

$$\mu(f) \geq 1.3248 \cdots > \theta = 1.3247 \ldots,$$

and proves thus Smyth's theorem.

1.10 Remarks

Parts of the proof of Smyth theorem can already be found in [Sieg], where it was proved that the real root of $x^3 - x - 1 = 0$ is the smallest Pisot number. The sharpest result in the direction of the general Lehmer conjecture is due to Dobrowolski, Cantor and Straus and Louboutin [Dobr], [Loub] which states that there exists a constant $\gamma > 0$ such that

$$H(\alpha)^n \geq 1 + \gamma \left(\frac{\log \log n}{\log n}\right)^3$$

for all $\alpha \neq 0$ of degree n which are not equal to a root of unity. The presentation chosen in this chapter, which led from the easy proof of Schinzel's and Zhang's theorem to Langevin's theorem, does not correspond to the correct chronological order of their discovery. Indeed, Langevin's theorem was published in 1985, and the former proofs were found almost ten years later. However, they are all three based on what is sometimes called the resultant method, which is already more or less explicitly used by Schinzel [Schi]. Zhang's theorem (along the lines of Zagier's proof) has been generalized by Beukers, Schieckewei, Schmidt, Wirsing, Zagier for obtaining absolutely lower bounds for heights along hypersurfaces; see [BeZa] and the references therein, and see the next section for a theorem of this kind. In particular, as corollary of the main result in [BeZa] one obtains a part of Smyth theorem: If the trace of α is integral and different from $1/\alpha$ (and α is thus not self-reciprocal), then $H(\alpha)^n \geq \sqrt{\frac{1+\sqrt{5}}{2}}$, where n is the degree of α. Another possible generalization of Zhang's theorem was investigated in [Smy2]. Here Zhang's theorem is interpreted as giving an absolute lower bound for the Mahler measure of polynomials in $X(X-1)$, and the paper generalizes this result to polynomials of the form $p(T(x))$, where $T(X) \in \mathbb{Z}[X]$ is of degree $n \geq 2$, divisible by X, but $\neq \pm X^n$. This point of view is also taken up in [Doch]

Part 2

Heights on Elliptic Curves

So far we have discussed heights of algebraic numbers. One may view this theory as theory of heights on the curve $\mathbb{P}^1$. Indeed, for a point $P = [x : y] \in \mathbb{P}^1(K)$, where K is a number field, define

$$H(P) = \prod_{v \in P_K} \max(|x|_v, |y|_v)^{1/[K:\mathbb{Q}]}.$$

By the product formula this does not depend on the choice of projective coordinates of P, and if we identify $\alpha \in K$ with the point $P := [\alpha : 1] \in \mathbb{P}^1(K)$, then $H(P) = H(\alpha)$. In this section we now discuss heights on curves of genus 1, which may be viewed as a natural step after the genus 0 case discussed before.

However, before going into this theory, we shall reinterprete Zhang's theorem. This theorem is in a sense on the boundary between the theory of heights of algebraic numbers and heights on general curves. Next, we have to discuss shortly heights on projective space, since some of the general results about such heights are needed for the theory of heights on elliptic curves.

2.1 Heights on affine plane curves

In this section we generalize the proof of Zhang's theorem as given in [Zagi]. For this we restate Zhang's theorem as a theorem about heights on affine, possibly reducible, plane algebraic curves defined over $\mathbb{Q}$. By such a curve we understand the set C of solutions (x, y) of an equation $F(x, y) = 0$, where $F \in \mathbb{Q}[x, y]$, and F is not constant. We use C^* for the curve defined by

$$F^*(x, y) := x^m y^n F(1/x, 1/y),$$

where m and n are the degrees of $F(x, y)$ in x and y respectively.

Zhang's theorem may be restated by saying that $H(\alpha)H(\beta) > C$ for all (α, β) on the curve $x + y = 1$. This suggests of thinking of $H(\alpha)H(\beta)$ as height of the point $P = (\alpha, \beta)$, and then Zhang's theorem says that the heights of the points on the line $x + y = 1$ are bounded to below. Or it may also be thought of saying that the heights of two algebraic numbers satisfying an algebraic (here linear) relation can not be both arbitrary small. It is not hard to generalize Zhang's theorem as follows:

Theorem 2.1. *Let C be an affine plane curve defined over $\mathbb{Q}$ such that C intersects C^* in only finitely many points. Then there is a constant $A > 1$ such that*

$$H(\alpha)H(\beta) \geq A$$

for all pairs of algebraic numbers (α, β) on C such that $\alpha, \beta \neq 0$ and (α, β) is not an intersection point of C with C^*.

Proof. Let $G(x, y)$ be a polynomial which vanishes at the intersection points of C with C^*. For real $s > 0$ set

$$\gamma_s(z, w) = |z|^{\frac{1}{2}}|w|^{\frac{1}{2}}|G(z, w)G(1/z, 1/w)|^s.$$

We show that for every sufficiently small $s > 0$ there is a constant $A = A_s > 1$ (depending on s) such that

$$\max(1, |z|)\max(1, |w|) \geq A_s\gamma_s(z, w)$$

for all (z, w) on the truncated curve $D := C \cup C^*$, which is defined by $FF^* = 0$, if, say C is defined by $F = 0$.

Since both sides of the desired inequality have the same invariance under $z \mapsto 1/z$ and under $w \mapsto 1/w$, it suffices to prove the estimate for all points on the curve $D_0 := D \cap (\mathbb{D} \times \mathbb{D})$, where $\mathbb{D}$ is the disk $|z| \leq 1$. Hence we have to show that for all $l \gg 0$

$$|z|^l|w|^l|G(z, w)G^*(z, w)| < 1$$

on D_0.

For proving this note that the number of points (z, w) of D_0 with $|zw| = 1$ is finite, and that, for any such point, one has $G(z, w) = 0$. Indeed, if (z, w) is such a point, then $(1/z, 1/w) = (\overline{z}, \overline{w})$, and hence, using that F and F^* have real coefficients, $F(z, w) = 0$ implies $F^*(z, w) = 0$ and vice versa, i.e. (z, w) is an intersection point of C and C^*. Hence, there is an open neighborhood U of all these points such that the last inequality holds true on U. Since $D_0 \setminus U$ is compact there exists a constant $R < 1$ such that $|zw| \leq R$ on

$D_0 \setminus U$. Moreover, $|GG^*| < a$ on $D_0 \setminus U$ with a suitable constant a. Thus, if l satisfies $R^{l'}a < 1$, where $l' = (l + \max(m, n))/2$ with m and n denoting the degree of G in x and y respectively, then the desired inequality holds true on all of D_0.

To finish the proof we proceed exactly as in the proof of Zhang's theorem. Let (α, β) is a pair of algebraic numbers on C, say of degree d and e and with normalized minimal polynomials $f = ax^d + \cdots$ and $g = bx^e + \cdots$, respectively. Then, for all sufficiently small s, we have

$$\begin{aligned} H(\alpha)^d H(\beta)^e &= |ab| \prod_{\alpha',\beta'} \max(1, |\alpha'|) \max(1, |\beta'|) \\ &\geq A_s^{d+e} |a|^{\frac{1}{2}-sm} |b|^{\frac{1}{2}-sn} |f(0)|^{\frac{1}{2}-sm^*} |g(0)|^{\frac{1}{2}-sn^*} \cdot \\ &\qquad \cdot \prod_{\alpha',\beta'} |a^{m+m^*} b^{n+n*} (GG^*)(\alpha', \beta')|^s, \end{aligned}$$

where m^* and n^* are the degrees of G^* in x and y, respectively, and where α' and β' are running through the conjugates of α and β. If

$$s \max(m, m^*, n, n^*) < 1/2$$

and if G has integral coefficients, then the right hand side is A^{d+e} times positive powers of nonnegative integers. Hence it is bounded to below by $\geq A^{m+n}$, unless $\alpha\beta(GG^*)(\alpha, \beta) = 0$.

We finally assume that we have chosen G such that the curves $D : GG^* = 0$ and C intersect in only finitely many points. If (α, β) is on C, but not on C^*, then α and β are not both roots of unity, and hence $H(\alpha)H(\beta) > 1$ by Kronecker's theorem. Thus, replacing A_s by the minimum of A_s and the $H(\alpha)H(\beta)$, where (α, β) runs through the finitely many points of C and $D : GG^* = 0$, but not on C^*, finally gives the desired estimate.

It remains to ensure the existence of a G with integral coefficients, vanishing on $C \cap C^*$, but such that $D : GG^* = 0$ and C have only finitely many points in common. Indeed, such polynomials exist. We can e.g. choose through each intersection P point of C and C^* a line $L_P(x, y) = 0$, such that neither this line, nor one of its finitely many conjugate lines $L_{P,j}(x, y) = 0$ lie on C or C^*. (A conjugate line is one whose defining equation is obtained by applying to the coefficients of L_P a Galois substitution of $\overline{\mathbb{Q}}$.) Then $G := \prod_{P,j} L_{P,j}$ has the desired properties. □

As already mentioned before, we had $C : x + y = 1$ in Zhang's theorem. Thus $C^* : x + y = xy$. The intersection points of C and C^* are $\rho = \frac{1+\sqrt{-3}}{2}$ and its complex conjugate. If we take for the G used in the preceding proof

$$G = (\rho x - \overline{\rho} y)(\overline{\rho} x - \rho y) = x^2 - xy + y^2,$$

then $G^* = x^2 - xy + y^2$, and

$$\gamma_s(1, 1-z) = |z|^{\frac{1}{2}}|1-z|^{\frac{1}{2}}\Big(\frac{|z^2-z+1|}{|z||1-z|}\Big)^{2s}.$$

This is the function we actually used in our original proof with $s = 1/4\sqrt{5}$.

2.2 Heights on projective space

For a point P in $\mathbb{P}^n$, say with projective coordinates $[x_0 : \cdots : x_n]$ in a number field K, we define its height $H_K(P)$ relative to K and its absolute height $H(P)$ by

$$H_K(P) = \prod_{v \in P_K} \max_{0 \le j \le n} |x_j|_v, \qquad H(P) = H_K(P)^{1/[K:\mathbb{Q}]}.$$

By the product formula $\prod_v |t|_v = 1$ $(t \in K)$ this is well defined (see the proof of 1.10), and by the compatibility relations $H(P)$ does not depend on the choice of the field K.

If $P \in \mathbb{P}^n(\mathbb{Q})$, then we may choose the projective coordinates x_j in $\mathbb{Z}$ and such that $\gcd(x_0, \ldots, x_n) = 1$. But then, for each non-archimedian v, we have $|x_j|_v \le 1$ for all j and $|x_j|_v = 1$ for at least one j, and hence $H(P)$ is given by the more intuitive formula

$$H(P) = \max_j |x_j|$$

with the usual archimedean absolute values $|x_j|$.

If $P = [x_0 : \cdots : x_n] \in \mathbb{P}^n(\mathbb{Q})$ and, say, $x_j \neq 0$, then the minimal field of definition $\mathbb{Q}(P)$ if P is defined as

$$\mathbb{Q}(P) = \mathbb{Q}\big(\frac{x_0}{x_j}, \ldots, \frac{x_n}{x_j}\big).$$

This does not depend on the choice of x_j.

We shall need two basic properties of the absolute height.

Theorem 2.2. *For each constant C and for each integer d the set*

$$\{P \in \mathbb{P}^n \,|\, H(P) \le C,\ [\mathbb{Q}(P) : \mathbb{Q}] \le d\}$$

is finite.

Proof. Indeed, one has for any $P \in \mathbb{P}^n$, say $P = [x_0 : \cdots : x_n]$ with at least one $x_j = 1$ and with $K = \mathbb{Q}(P)$,

$$H_K(P) = \prod_{v \in P_K} \max_j |x_j|_v \geq \max_j \prod_v \max(1, |x_j|_v) = \max_j H_K(x_j).$$

If $[\mathbb{Q}(P) : \mathbb{Q}] \leq d$ then we also have $[\mathbb{Q}(x_j) : \mathbb{Q}] \leq d$ for all j. Thus the theorem follows from the special case $n = 1$, which we proved in section 1.4. □

By a morphism

$$F : \mathbb{P}^n \longrightarrow \mathbb{P}^m$$

of degree d we understand a map of the form

$$F(P) = [f_0(x_0, \ldots, x_n), \ldots, f_n(x_0, \ldots, x_n)], \qquad (P = [x_0 : \cdots : x_n]),$$

where the f_j are homogeneous polynomials of degree d and with coefficients in $\overline{\mathbb{Q}}$. In particular, for such a set of polynomials f_j, one has $f_j(x_0, \ldots, x_n) = 0$ for all $0 \leq j \leq m$ if and only if $x_0 = x_1 = \cdots = x_m = 0$.

Theorem 2.3. *Let $F : \mathbb{P}^n \longrightarrow \mathbb{P}^m$ be a morphism of degree d. Then there exist constants $C_1, C_2 > 0$ such that*

$$C_1 H(P)^d \leq H(F(P)) \leq C_2 H(P)^d$$

for all $P \in \mathbb{P}^n$.

Proof. Let $P = [x_0 : \cdots : x_n] \in \mathbb{P}^n(K)$. For a place $v \in P_K$ we set $\varepsilon(v) = 1$ if v is archimedean, and $\varepsilon(v) = 0$ otherwise. Using this symbol we have, for all v and all points $a_j \in K$,

$$|a_1 + \cdots + a_r|_v \leq r^{\varepsilon(v)} \max_{1 \leq j \leq r} |a_j|_v.$$

Moreover, we use $H_v(P) = \max_j |x_j|_v$, thus $H_K = \prod_v H_v$.

Accordingly we have (using $f_{j,k}$ for the $\binom{n+d}{d}$ coefficients of f_j)

$$H_v(F(P)) = \max_j |f_j([x_0 : \cdots : x_n])|_v$$
$$\leq \binom{n+d}{d}^{\varepsilon(v)} \big(\max_{j,k} |f_{j,k}|_v\big)\big(\max_j |x_j|_v^d\big).$$

This yields the second inequality.

For the first one we need the Hilbert Nullstellensatz (see any text book an algebraic geometry). In our case it asserts that, for any polynomial f which vanishes at the common zeroes of all the f_j, some positive integral power f^r lies in the ideal I generated by the f_j in the ring $\overline{\mathbb{Q}}[X_0, \ldots, X_n]$. Now, the is the only common zero of the f_j is the point 0, and hence, for a suitable integer r the polynomials $X_0^r, \ldots, X_n^r$ lie in I. In other words $X^r = \sum_j P_{k,j} f_j$ with suitable $P_{k,j} \in \overline{\mathbb{Q}}[X_0, \ldots, X_n]$. These identities remain valid if we replace the $P_{k,j}$ by their $r - d$th homogeneous component, and hence we may assume that the $P_{k,j}$ are homogeneous of degree $r - d$. Enlarging K if necessary, we may furthermore assume that the $P_{k,j}$ have coefficients in K. Then, similar to the reasoning above, we have

$$\begin{aligned} H_v(P)^r &= \max_k \Big| \sum_j (P_{k,j} f_j)(x_0, \ldots, x_n) \Big|_v \\ &\le (m+1)^{\varepsilon(v)} \big(\max_{k,j} |P_{k,j}(x_0, \ldots, x_n)|_v \big) \big(\max_j |f_j(x_0, \ldots, x_n)|_v \big) \\ &\le (m+1)^{\varepsilon(v)} \binom{n + r - d}{r - d}^{\varepsilon(v)} C\, H_v(P)^{r-d} H_v(F(P)), \end{aligned}$$

where C is the maximum of the v-adic valuations of the coefficients of all the $P_{k,j}$. This implies the first estimate. □

2.3 Plane curves as diophantine equations

Everybody knows how to compute $L(\mathbb{Q})$ for a line $L/\mathbb{Q}$ in the projective plane $\mathbb{P}^2$. It is also not difficult to compute $C(\mathbb{Q})$ for a projective plane curve $C/\mathbb{Q}$ of degree 2. Let us consider, to have a concrete example, the circle C which is given in affine coordinates by $C : x^2 + y^2 = 1$. We fix a point $O \in C(\mathbb{Q})$. Then, for $P \in C(\mathbb{Q})$, the line L_P through O and P is defined over $\mathbb{Q}$. If $P = (x_1, y_1)$, then L_P is given by the equation

$$y = \frac{y_1}{x_1 - 1}(x - 1).$$

Conversely, if L is a line through O, then L intersects C in exactly two points, in O and in a second point $P = (x_1, y_1)$. (If $P = O$ then L_P is the tangent to C at O and vice versa.) If L is defined over $\mathbb{Q}$, then so is P. Indeed, if L is given by $y = \lambda(x - 1)$, then x_1 is a solution of the quadratic equation over $\mathbb{Q}$ obtained by replacing y in $x^2 + y^2 = 1$ by $\lambda x + \mu$. Since $x = 1$ is also a solution, x_1 is necessarily rational, and so is $y_1 = \lambda x_1 + \mu$. Working out the details one finds $x_1^2 + \lambda^2(x_1 - 1)^2 = 1$, i.e.

$$x_1 = \frac{\lambda^2 - 1}{\lambda^2 + 1}, \quad y_1 = \frac{-2\lambda}{\lambda^2 + 1}.$$

In general, if $C/\mathbb{Q}$ is an irreducible smooth projective plane curve of degree 2, and $O = (x_0, y_0) \in C(\mathbb{Q})$, then one can easily verify that the map

$$C \to \mathbb{P}^1, \quad P = (x_1, y_1) \mapsto \frac{y_1 - y_0}{x_1 - x_0} = \text{slope of the line through } O \text{ and } P$$

is an isomorphism defined over $\mathbb{Q}$ and mapping $C(\mathbb{Q})$ onto $\mathbb{P}^1(\mathbb{Q})$. The above method of determining $C(\mathbb{Q})$ is effective, apart from the fact that we have to find at least one $O \in C(\mathbb{Q})$ to start with.

We now turn to cubic curves defined over $\mathbb{Q}$. Here the situation has still some similarities with the quadratic case, though there are also much more complications. Again we start with the idea of reducing to algebraic equations in one variable by intersecting with lines. However, if we intersect a cubic curve $C/\mathbb{Q}$ with a line, then there will be in general three intersection points. But still, if the line is defined over $\mathbb{Q}$ and two of the intersection points are in $C(\mathbb{Q})$, then the third one belongs to $C(\mathbb{Q})$ too. However, one can make an even stronger statement.

To explain this we restrict for the following to elliptic curves in Weierstrass form defined over a number field K. By such a curve we understand a cubic curve E which, in affine coordinates, is given by an equation of the form

$$E : y^2 = x^3 + Ax + B,$$

where A, B are elements of K, and where we assume that the the polynomial in x on the right has no multiple roots, i.e. that its discriminant

$$\Delta_E := -4A^3 - 27B^2 \neq 0.$$

Such a curve has exactly one point O on the line at infinity, which in homogeneous coordinates is given by

$$O = [0 : 1 : 0].$$

The condition $\Delta_E \neq 0$ ensures that E is a non-singular curve. The restriction to such curves is not a serious one, since any non-singular plane cubic curve is isomorphic to a curve in Weierstrass form (see the next section for details).

If for $P = (x, y) \in E$ we set $-P := (x, -y)$, and if we define a binary operation $+$ on E by letting $P_1 + P_2$ the unique point P such that P_1, P_2 and $-P$ (counting multiplicities) are the intersection points of E with a line, then E becomes a group (for details and a proof of this see the next section). Clearly, the point 0 at infinity is the neutral element of E (it is an inflection point), and if α is a root of $X^3 + AX + B$, then $(\alpha, 0)$ is a point of order 2. Finally, if E is defined over K, then $E(K)$ is a subgroup of E. This follows easily by looking at the equations expressing the affine coordinates of $P_1 + P_2$ in terms of those of P_1, P_2 (again, see the next section for details).

Assume now, to come back to diophantine equations over $\mathbb{Q}$ and to show the idea for the general theory developed in a moment, that E is of the special form

$$E : y^2 = (x-a)(x-b)(x-c)$$

with pairwise different integers a, b, c. Clearly the question is when, for a rational number x, the product $(x-a)(x-b)(x-c)$ is a square in $\mathbb{Q}$. To analyze this we introduce the map

$$\phi : E(\mathbb{Q}) \to G := (\mathbb{Q}^*/\mathbb{Q}^{*2})^2,$$

$$P \mapsto \begin{cases} \big((x-a)\mathbb{Q}^{*2}, (x-b)\mathbb{Q}^{*2}\big) & \text{if } x \neq a, b,\ P \neq 0 \\ 1 & \text{if } P = 0 \\ \big((x-b)(x-c)\mathbb{Q}^{*2}, (x-b)\mathbb{Q}^{*2}\big) & \text{if } x = a \\ \big((x-a)\mathbb{Q}^{*2}, (x-a)(x-c)\mathbb{Q}^{*2}\big) & \text{if } x = b \end{cases} ,$$

where (x, y) are the affine coordinates of P if $P \neq 0$.

Lemma 2.1. *The map ϕ is a group homomorphism with kernel $2E(\mathbb{Q})$.*

Proof. For showing that ϕ is a group homomorphism it clearly suffices to show that $\phi(P_1)\phi(P_2)\phi(P_3) = 1$ if $P_1 + P_2 + P_3 = 0$. This is trivial if one of he P_j is 0. Otherwise the P_j lie on a line $y = \lambda x + \mu$ with $\lambda, \mu \in\in \mathbb{Q}$, $\lambda \neq 0$. Hence, if we set $P_j = (x_j, y_j)$, then the x_j are the solutions of

$$(x-a)(x-b)(x-c) - (\lambda x + \mu)^2 = 0,$$

Hence we have

$$(x-a)(x-b)(x-c) - (\lambda x + \mu)^2 = (x - x_1)(x - x_2)(x - x_3).$$

In particular, considering this equation for $x = a$, $x = b$ and $x = c$, respectively, we observe that

$$(a-x_1)(a-x_2)(a-x_3), (b-x_1)(b-x_2)(b-x_3), (c-x_1)(c-x_2)(c-x_3) \in \mathbb{Q}^2.$$

From this one easily obtains $\phi(P_1)\phi(P_2)\phi(P_3) = 1$.

Clearly $2E(\mathbb{Q})$ is mapped to 1 since $\mathbb{Q}^*/\mathbb{Q}^{*2}$ has exponent 2. Conversely, assume that $\phi(P) = 1$. . . . to be completed later. □

For $v \in P_{\mathbb{Q}}$, let G_v denote the subgroup (of order 2) in $\mathbb{Q}^*/\mathbb{Q}^*$ generated by $p\mathbb{Q}^{*2}$ if v is non-archimedean belonging to the prime number p, and by $(-1)\mathbb{Q}^{*2}$, if v is archimedean. Clearly, $\mathbb{Q}^*/\mathbb{Q}^{*2} = \sum_{v \in P_{\mathbb{Q}}} G_v$.

Lemma 2.2. *The image of ϕ is contained in*

$$\bigoplus_{p|\Delta_E \text{ or } p=\infty} G_p^2,$$

where $\Delta_E = (a-b)^2(a-c)^2(b-c)^2$ is the discriminant of E. In particular, it is finite.

Proof. Let $P \in E(\mathbb{Q})$, $P \neq 0$, say $P = (x, y)$. Let p be a prime number, and let $\phi(P) = (u\mathbb{Q}^{*2}, v\mathbb{Q}^{*2})$. We have to show that $\mathrm{ord}_p(u)$ and $\mathrm{ord}_p(v)$ are both even.

For this let p^n be the exact power of p in the prime decomposition of x.

If $n < 0$, then $x \neq 0, a, b$ and we can take $u = x - a$ and $v = x - b$. Since a, b, c are integral p^n is also the exact power of p in $x - a$, $x - b$ and $x - c$. We have accordingly $\mathrm{ord}_p(y^2) = 3n$. On the other hand $\mathrm{ord}_p(y^2)$ is even. It follows that $n = \mathrm{ord}_p(u) = \mathrm{ord}_p(v)$ is even.

If $n \geq 0$, then the order at p of each of the three numbers $x - a$, $x - b$ and $x - c$ is nonnegative. At most one of them has positive order since the difference of two of any of these divides Δ. Again, since their product is a square in $\mathbb{Q}$, this implies that the orders at p of these numbers are even. Hence if $x \neq a, b$ then u and v have even order.

The case $n \geq 0$ and $x = a$ or $x = b$ is left to the reader. □

Let R be a set of representatives for $E(\mathbb{Q})/2E(\mathbb{Q})$. By the preceding lemma R is a finite set. The set R (and possibly a finite number of additional points in $E(\mathbb{Q})$ to be explained in a moment) play the role of the point O in the case of quadrics considered above. Namely, let $P_0 \in E(\mathbb{Q})$. Then we can find an $Q \in R$ such that $P_0 = Q_0 + 2P_1$ for some $P_1 \in E(\mathbb{Q})$. Again, we find a $Q_1 \in R$ such that $P_1 = Q_1 + 2P_2$ with a suitable $P_2 \in E(\mathbb{Q})$, and so forth. Suppose that in each step the point P_j is of less complexity, say needs less digits to be described, than its predecessor P_{j-1}. Then we may hope that our descent procedure will end in the sense that P_n for some n is in a finite set S of very simple points. Hence P_0 is a linear combination of the points in $R \cup S$, which solves the problem of determining $E(\mathbb{Q})$. That the group $E(\mathbb{Q})$ is finitely generated is indeed the case for any elliptic curve over $\mathbb{Q}$; this is Mordell's theorem which we shall prove in the next sections following exactly the ideas sketched in this paragraph. The complexity of points in $E(\mathbb{Q})$ will of course be measured using a height function.

For curves of genus strictly greater than 1 the situation is completely different from the genus 0 and 1 case. Here one has Mordell's conjecture, which was proved by Faltings (for another proof, based on Faltings', but shorter, more self contained and using the theory of heights instead of arithmetic intersection theory, see [Bomb]).

Theorem 2.4. *(Mordell-Faltings) For a projective curve $C/\mathbb{Q}$ with genus ≥ 2 the set $C(\mathbb{Q})$ of its rational points is finite.*

Thus, for curves of genus 2 the problem is too find good a priori upper bounds for the height (to be properly defined in some sense) of its rational points.

2.4 Basic facts about elliptic curves

This section is still incomplete. To complete the logical thread of this second part the following topics would have to be reviewed: group law — E(K) — K(P)— Weierstrass form — action of Galois [m] is surjective — affine and projective form — K(E) = maps onto $\mathbb{P}^1$ — deg(f) — $\widetilde{E}$ — E as Jacobian of itself

2.5 Heights on elliptic curves

We fix for this section an elliptic curve E defined over a number field K, which we suppose always to be given in Weierstrass form

$$E : y^2 = x^3 + Ax + B, \qquad (A, B \in K)$$

As height $H_0(P)$ of a point $P \in E$, say with homogeneous coordinates $[x : y : z]$ in a number field L, we may consider the height of P considered as point of the projective plane $\mathbb{P}^2$, i.e.

$$H_0(P) = \prod_{v \in P_L} \max(|x|_v, |y|_v, |z|_v)^{1/[L:\mathbb{Q}]}.$$

Another possibility would be to view x as a function from E onto $\mathbb{P}^1$, and to take $H_x(P) := H(x(P))$ as the height of P, where $H(x(P))$ is the height of $x(P)$ as point of $\mathbb{P}^1$. Or, more generally, we could take any nonconstant function $f \in K(E)$, consider it as function onto $\mathbb{P}^1$ and take $H_f(P) := H(f(P)$ as height function.

However, as it turns out, all these possibilities are essentially equivalent. Also, notations become more natural if one uses additive notation, i.e. if one uses the logarithmic heights

$$h_f(P) := \frac{1}{\deg f} \log H(f(P)).$$

The reason for normalizing be the factor $1/\deg f$ will become clear in a moment.

Theorem 2.5. *Let $f, g \in K(E)$ be nonconstant functions on E. Then, for every $\varepsilon > 0$, there are constants $C_1, C_2 > 0$ such that*

$$C_1 H_f(P)^{-\varepsilon} \leq \frac{H_f(P)^{1/\deg f}}{H_g(P)^{1/\deg g}} \leq C_2 H_f(P)^{+\varepsilon}$$

for all P. Or, using logarithmic heights, for every $\varepsilon > 0$, there is a constant C such that

$$|h_f(P) - h_g(P)| \leq C + \varepsilon h_f(P)$$

for all $P \in E$.

Proof. It is easy to check that the last inequality defines an equivalence relation on the set of all functions h_f with f running through the non constant elements of $K(E)$. Hence it suffice to prove the last inequality for some specific choice of g and arbitrary f. We choose $g = x$. Moreover, we assume also that f is even. For the general case we refer the reader to [Weil] (or [Lan1], Ch. 4, Cor. 3.5). Here we call f even if $f(-P) = f(P)$. For even f the desired inequality is in fact true even for $\varepsilon = 0$.

Now f is a rational function in x and y, say $f = p(x, y)/q(x, y)$, with two polynomials $p, q \in \overline{\mathbb{Q}}[X, Y]$. Since y^2 is a polynomial in x we can even write $f = (p_1(x) + yp_2(x))/(q_1(x) + yq_2(x))$ with polynomials $p_j, q_j \in \overline{\mathbb{Q}}[X]$. Also, we may assume that the numerator and denominator are relatively prime. Then they are unique up to multiplication by constants. But then we observe, on using that y is an odd function, i.e. $y(-P) = -y(P)$, that f can only be even if $p_2 = q_2 = 0$.

Hence $f = r \circ x$, where r is the rational function $r : \mathbb{P}^1 \to \mathbb{P}^1$ given by $r(t) = p_1(t)/q_2(t)$. Since any such rational function is a morphism (in the sense explained in section 2.2), the theorem for f and $g = x$ now follows from Theorem 2.3: there exists constants $C_1, C_2 > 0$ such that

$$C_1 H(x(P))^{\deg r} \leq H((r \circ x)(P))^{\deg r} \leq C_2 H(x(P))^{\deg r}.$$

Using $\deg f = 2 \deg r$ we obtain the desired inequality. □

The heights h_f possess a striking property, which we shall use to derive a canonical height from them by a procedure analogous to the one which led us to the definition of the Mahler measure.

Theorem 2.6. *Let $f \in K(E)$. Then there is a constant C such that*

$$|h_f(P + Q) + h_f(P - Q) - (2h_f(P) + 2h_f(Q))| \leq C$$

for all $P, Q \in E$.

Proof. It suffices to prove this identity for some particular function f. The general result follows then from the preceding theorem. For f we choose the coordinate function x.

For the proof we look at the following diagram:

$$\begin{tikzcd}
E \times E \arrow[r, "\phi"] \arrow[d, "x \times x"'] & E \times E \arrow[d, "x \times x"'] \\
\mathbb{P}^1 \times \mathbb{P}^1 \arrow[r] \arrow[d, "\iota"'] & \mathbb{P}^1 \times \mathbb{P}^1 \arrow[d, "\iota"'] \\
\mathbb{P}^2 \arrow[r, "\underline{\phi}"] & \mathbb{P}^2
\end{tikzcd}$$

Here we use

$$\phi : (P, Q) \mapsto (P + Q, P - Q),$$
$$\iota : ([x : y], [x'y']) \mapsto [yy', xy' + x'y, xx''],$$
$$\underline{\phi} : [a : b : c] \mapsto [b^2 - 4ac : 2b(Aa + c) + 4Ba^2 : (c - Aa)^2 - 4Bab].$$

(Here A, B are the coefficients of the Weierstrass equation defining E.) It is not completely obvious, though straightforward, to check that the diagram is commutative and that $\underline{\phi}$ is a morphism (see section 2.2).

Moreover, we leave it as an exercise to verify that there exist constants $C_1, C_2 > 0$ such that

$$C_1 \leq \frac{H(A)H(B)}{H(\iota(A, B))} \leq C_2$$

for all $P, Q \in \mathbb{P}^1$.

We use $h(A) := \log H(A)$ for $A \in \mathbb{P}^n$ and H denoting the height on $\mathbb{P}^n$. Finally, for any two real valued functions α, β on $E \times E$ we write $\alpha \approx \beta$ if $|\alpha\beta|$ is bounded on $E \times E$. We then have

$$\begin{aligned}
h_x(P + Q) + h_x(P - Q) &= h(x(P + Q)) + h(x(P - Q)) \\
&\approx h(\iota(x(P + Q), x(P - Q))) \\
&= h(\underline{\phi} \circ i(x(P), x(Q))) \\
&\approx 2h(\iota(x(P), y(Q))) \approx 2h(x(P)) + 2h(x(Q)).
\end{aligned}$$

Here, for the last but not least identity we used Theorem 2.3 and that the degree of $\underline{\phi}$ is 2. This proves the desired estimates. □

We now define the canonical height (or Néron-Tate) height of a point P on E by

$$h(P) = \lim_k \frac{1}{4^k} h_f(2^k P).$$

If we right n for 2^k and if we use that $E[n]$ consists of exactly n^2 points, then $h(P)$ can be viewed more suggestively as the limit of the sequence

$$\frac{1}{n^2}h_f(\sum_{\substack{Q\in E\\ nQ=P}} Q).$$

This is exactly the kind of formula (written additively) which we used to define the Mahler measure. In fact, it could be shown that, instead of powers of 2, we can take powers of any arbitrary nonnegative integer for obtaining the same limit.

Theorem 2.7. *The limit defining $h(P)$ converges uniformly in P. It does not depend on the choice of f. There is a constant C such that*

$$|h(P) - h_f(P)| \leq C$$

for all $P \in E$.

Proof. By the last theorem, setting $Q = P$, we obtain that

$$|h_f(2P) - 4h_f(P)| \leq C$$

for all P with a constant independent of P. We use this to show that $4^{-k}h(2^kP)$ is a Cauchy sequence uniformly in P. Indeed, if $m > n$ then, using the above estimate, we obtain

$$|4^{-m}h_f(2^mP) - 4^{-n}h_f(2^nP)| = \sum_{k=n}^{m-1} |4^{-(k+1)}h_f(2^{k+1}P) - 4^{-k}h_f(2^kP)|$$

$$\leq \sum_{k=n}^{m-1} \frac{C}{4^{k+1}} < \frac{4C}{3\cdot 4^{n+1}},$$

which tends to zero, independent of P, for $m, n \to \infty$.

The last assertion of the theorem follows similarly by writing

$$h(P) - h_f(P) = \sum_{k=0}^{\infty} 4^{-(k+1)}h_f(2^{k+1}P) - 4^{-k}h_f(2^kP).$$

If g is another nonconstant function on E, then, for each $\varepsilon > 0$, we have $|h_f(P) - h_g(P)| \leq \varepsilon h_f(P) + C$ with a constant independent of P. Replacing here P by 2^kP, dividing by 4^k and letting k tend to infinity shows that the difference of the limits of $4^{-k}h_g(2^kP)$ and $4^{-k}h_f(2^kP)$ is bounded by ε times the second limit. Since this is true for all $\varepsilon > 0$ the two limits must be equal. □

Immediately from the definition we obtain that h is an even function and that $h(0) = 0$, as follows easily on taking x for f in the definition of h. Similarly, one obtains

Theorem 2.8. *For each $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/K)$ and each $P \in E$ one has $h(P^\sigma) = h(P)$.*

Proof. This follows on writing $h(P)$ as limit of $\log H(x(nP))^{1/n}$ $(n = 2^k)$, and using $H(x(P^\sigma)) = H(x(P)^\sigma) = H(x(P))$, where the last identity is obvious from the very definition of the height H on $\mathbb{P}^2$. □

Theorem 2.9. *For each constant C and each integer d, the set*

$$\{P \in E \,|\, h(P) \leq C,\ [\mathbb{Q}(P) : \mathbb{Q}] \leq d\}$$

is finite.

Proof. Since $h_x(P) \leq h(P) + C$ for some constant C it suffices to prove the theorem with h replaced by h_x. But this is an immediate consequence of the fact that the map $P \mapsto x(P)$ is two-to-one, and the fact that there is only a finite number of algebraic numbers with height and degree below a fixed bound (see section 1.4). □

An important property is that the height is a quadratic form as is already suggested by the quasi-parallelogram law for the h_f as stated in Theorem 2.6

Theorem 2.10. *The map*

$$\langle\ ,\ \rangle : E \times E \to \mathbb{R},\ \langle P, Q\rangle := h(P+Q) - h(P) - h(Q)$$

is $\mathbb{Z}$-bilinear. In particular, one has $h(nP) = n^2 h(P)$ for all integers n and all P.

Proof. By writing in Theorem 2.6 nP and nQ for P and Q, dividing by n^2 and letting n tend to infinity we obtain the so-called parallelogram law

$$h(P+Q) + h(P-Q) = 2h(P) + 2h(Q).$$

From this the bilinearity follows by a simple algebraic manipulation. Since the pairing $\langle\ ,\ \rangle$ is symmetric it suffices to prove

$$\langle P+Q, R\rangle = \langle P, R\rangle + \langle Q, R\rangle.$$

It is straightforward to check that this is equivalent to

$$h(P+Q+R) - h(P+Q) - h(P+R+) - h(Q+R) + h(P) + h(Q) + h(R) = 0.$$

But this follows indeed from the parallelogram law (and using the evenness of h) as follows. Applying four times the parallelogram law gives

$$\begin{aligned}
&h(P+Q+R)+h(P+Q-R)-2h(P+Q)-2h(R)=0\\
&h(P-Q+R)+h(P+Q-R)-2h(Q-R)-2h(P)=0\\
&h(P-Q+R)+h(P+Q+R)-2h(P+R)-2h(Q)=0\\
&2h(Q+R)+2h(Q-R)-2h(Q)-2h(R)=0,
\end{aligned}$$

and taking the alternate sum of these four equations is exactly the desired identity..

The second assertion follows from $2h(P) = h(P) + h(-P) - h(P-P) = -\langle P, -P\rangle = \langle P, P\rangle$. □

As direct generalization of Kronecker's theorem one has

Theorem 2.11. *One has $h(P) = 0$ if and only if P is a torsion point.*

Proof. By the preceding theorem we clearly have $\langle P, P\rangle = 0$ if $nP = 0$ for some integer $n \geq 1$. Conversely, if $h(P) = 0$, then $h(nP) = 0$ for all P. If L/K is a number field such that $P \in E(L)$, then $nP \in E(L)$ for all n. But the set of all $Q \in E(L)$ with $h(Q) = 0$ is finite as we saw above. Hence P must have finite order. □

Since the set of points on E with height below a given bound affine coordinates in a given number field L is finite, we see that in particular $E(K)_{\mathrm{tor}}$ is finite. However, one can say much more. The theorem of Mazur [] says that, for an E defined over $\mathbb{Q}$ the subgroup $E(\mathbb{Q})_{\mathrm{tor}}$ is always isomorphic to one of a given list of fifteen abelian groups. It is conjectured that this is true for all number fields K in the following sense: For each number field K there is a constant N such that $E(K)_{\mathrm{tor}}$, for any elliptic curve E defined over K, has not more than N points. By a theorem of Manin [Mani] one knows at least that for any K and any prime number p there exists a constant N such that the p-part of $E(K)_{\mathrm{tor}}$, for any E/K, is bounded to above by N.

From the last theorem we also obtain

Theorem 2.12. *The height pairing $\langle\ ,\ \rangle$ on E factors to a non-degenerate pairing $E/E_{tor} \times E/E_{tor} \to \mathbb{R}$.*

Proof. Clearly $\langle P, Q\rangle = 0$ for all Q if $nP = 0$ for some $n \geq 1$. Conversely $\langle P, Q\rangle = 0$ for all Q implies $h(P) = 0$, and hence that P is a torsion point. □

We conclude this section with another result showing that the canonical height deserves its name.

Theorem 2.13. *Let h' be a real valued function on E which satisfies the two following properties:*

1. *There exists an integer $n \geq 2$ such that $h'(nP) = n^2 h'(P)$ for all $P \in E$.*

2. *There exists a function $f \in E(K)$ and a constant C such that $|h(P) - h_f(P)| \leq C$ for all $P \in E$.*

Then $h' = h$.

Proof. From the second assumption we see that $|h'(P) - h(P)| \leq C'$ for all P with a suitable constant C' (not depending on P). But then from the first assumption $h'(n^k P) = n^{2k} h'(P)$ for all $k \geq 0$, and hence

$$|h'(P) - h(P)| = \frac{1}{n^{2k}} |h'(n^k P) - h(n^k P)| =\leq \frac{C'}{n^{2k}}$$

for all k, whence, for $k \to \infty$, we obtain $h'(P) = h(P)$. □

2.6 Infinite descent on elliptic curves

In this section, using the theory of heights on elliptic curves, we can finally make precise the infinite descent procedure described at the the end of section 2.3. For this let E be a given elliptic curve defined over the number field K. We shall prove in the next section, that $E(K)/mE(K)$ is a finite group for each integer $m \geq 2$. As already indicated before this, together with the infinite descent procedure, implies that $E(K)$ is a finitely generated group. The descent procedure is effective, i.e. it shows how to calculate generators for $E(K)$ (provided we we can compute a set of representatives for the quotient $E(K)/mE(K)$).

Let $\mathfrak{R}$ be a system of representatives for $E(K)/mE(K)$ for a fixed $m \geq 1$. For this set of representatives let

$$C := 2 \max\{h(P) \,|\, P \in \mathfrak{R}\}.$$

We then have, for all $P \in E$ and all $R \in \mathfrak{R}$.

$$h(P + R) = 2h(P) - h(P - R) + 2h(R) \leq 2h(P) + C.$$

Let now $P \in E(K)$. We define a sequence of points $P_l \in E$ and $Q_l \in \mathfrak{R}$ by $P_0 = P$ and for $l \geq 1$

$$mP_l = P_{l-1} - Q_{l-1}.$$

Then

$$\begin{aligned}
h(P_l) &\leq \frac{1}{m^2}(2h(P_{l-1}) + C) \\
&\leq \frac{2^l}{m^{2l}}h(P) + C(\frac{1}{m^2} + \frac{2}{m^4} + \frac{4}{m^6} + \cdots + \frac{2^{l-1}}{m^{2(l-1)}}) \\
&\leq \frac{2^l}{n^{2l}}h(P) + \frac{C}{m^2 - 2}.
\end{aligned}$$

Finally, let $\mathfrak{R}_0$ be the set of all $Q \in E(K)$ with $h(Q) \leq C/(m^2 - 2)$. This is a finite set. Since the set of all Q with $h(Q) \leq C/(m^2 - 2) + .1$ is also finite, we can find a $\varepsilon > 0$ such that $\mathfrak{R}_0$ coincides with the set of all $Q \in E(K)$ with $h(Q) \leq C/(m^2 - 2) + \varepsilon$.

But then we conclude that $P_l \in \mathfrak{R}_0$, if l is large enough. In other words the set $\mathfrak{R} \cup \mathfrak{R}_0$ is a set of generators for $E(K)$. The set $\mathfrak{R}_0$ can be calculated by a systematic search.

2.7 The Mordell-Weil theorem

Again, throughout this section, E denotes an elliptic curve defined over a number field K. Moreover we fix an integer $m > 0$. The purpose of this section is to prove

Theorem 2.14. *(Weak Mordell-Weil theorem) The group $E(K)/mE(K)$ is finite.*

Together with the infinite descent procedure of the last section this implies then strong Mordell-Weil theorem

Theorem 2.15. *The group $E(K)$ is finitely generated.*

The proof of the so-called weak Mordell-Weil theorem has actually nothing to do with heights, but uses what is called Kummer theory for elliptic curves. However, we include it here for the sake of completeness. The Mordell-Weil theorem was actually first proved by Mordell for the case of an elliptic curve over $\mathbb{Q}$, was before already conjectured by Poincaré, and later generalized to arbitrary K (and arbitrary abelian varieties) by Weil, based on work of Siegel who introduced the powerful tool of heights into the study of diophantine problems. The proof uses the two fundamental finiteness theorem of algebraic number theory, the finiteness of class numbers and Dirichlet's unit theorem.

We shall show first that we can enlarge K without restriction of generality.

Lemma 2.3. *Let L/K be a finite extension. If $E(L)/mE(L)$ is finite, then so is $E(K)/mE(K)$.*

Proof. Let N be the kernel of the natural map

$$E(K)/mE(K) \to E(L)/mE(L);$$

thus $N = (E(K) \cap mE(L))/mE(K)$. We have to show that N is finite.

For each $C \in N$ pick a $P \in C$, and then a $Q \in E$ such that $P = mQ$. we set

$$\lambda_C : \mathrm{Gal}(L/K) \to E[m], \quad \lambda_C(\sigma) = Q^\sigma - Q.$$

Note that indeed $\lambda_C(\sigma) \in E[m]$ since $mQ^\sigma = P^\sigma = P = mQ$. If $\lambda_C = \lambda_{C'}$, say $C' = P' + mE(K)$ with associated $mQ' = P'$, then $Q - Q'$ is invariant under all $\sigma \in \mathrm{Gal}(L/K)$, and is hence in $E(K)$. But this means $P - P' \in mE(K)$, i.e. $C = C'$. Thus the map $C \mapsto \lambda_C$ is injective; its image being finite implies the lemma. □

The proof, being a little bit puzzling at the first glance, has a very natural explication in term of Galois cohomology. We shall explain this below (see section 2.8).

In the following we can hence assume, by enlarging K if necessary, that

$$E[m] \subset E(K).$$

Note that this implies in particular the following: If $Q \in E$ is such that $mQ \in E(K)$, then $L := K(Q)$ is a Galois extension of K. Indeed, if $\sigma : L \to \mathbb{C}$ is an embedding leaving K invariant, then $L^\sigma = K(Q^\sigma)$. But $Q^\sigma \in Q + E[m]$ (since $m(Q^\sigma) = (mQ)^\sigma) = mQ$), and hence $Q^\sigma \in L$, i.e. $L^\sigma = L$.

We set

$$L := K(Q \,|\, QP \in E(K))/qquadG := \mathrm{Gal}(L/K).$$

Then L is a Galois extension of K (a priori possibly infinite). We have a map

$$E(K) \times G \to E[m],$$

given by

$$(P, \sigma) \mapsto Q^\sigma - Q,$$

where Q is any point of E such that $mQ = P$. (We recall that such a point Q always exists since multiplication by m is a nonconstant morphism.)

Note that this definition does not depend on a particular choice of Q since any two inverse images of P under multiplication by m differ by an element of $E[m]$, which, as subset of $E(K)$, is invariant under G. The map is actually bilinear. It is linear in the right argument since

$$Q^{\sigma\tau} = (Q^\sigma - Q)^\tau + (Q^\tau - Q) = (Q^\sigma - Q) + (Q^\tau - Q),$$

where we used that $Q^\sigma - Q$ is in $E[m]$ and hence stable under G. It is obviously linear in the first argument.

The left kernel of the pairing (i.e. the subgroup of $P \in E(K)$ such that $\langle P, G\rangle = 0$) clearly contains $mE(K)$; in fact, it equals $mE(K)$. Indeed, if a $Q \in E$ with $P := mQ \in E(K)$ satisfies $Q^\sigma = Q$ for all $\sigma \in G$, then $Q \in E(K)$, i.e. $P = mQ \in mE(K)$. Thus the above pairing factors through a pairing

$$E(K)/mE(K) \times \mathrm{Gal}(L/K) \to E[m],$$

the so-called Kummer pairing, which is left non-degenerate. Or, to state this differently, the associated homomorphism

$$E(K)/mE(K) \to \mathrm{Hom}(G, E[m])$$

is injective. For proving the weak Mordell theorem it thus suffices to show that L is a finite extension of K. Hence, we start now to investigate more closely the field L.

First of all we note that the Kummer pairing is even perfect. Namely, for a fixed σ, let $Q^\sigma = Q$ for all Q with $mQ \in E(K)$. This means that σ leaves invariant L, and hence equals 1. Hence G embeds injectively into $\mathrm{Hom}(E(K)/mE(K), E[m])$, In particular, L is abelien with exponent m.

We now assume that E is given by a Weierstrass equation of the form $y^2 = x^3+Ax+B$ with A and B being integral algebraic integers (in K). This is no restriction of generality since for each pair $A, B \in K$ we can find an integer $N > 0$ such that N^4A and N^6B are integral, and we may the consider $y^2 = x^3 + N^4AX + N^6B$, which is isomorphic to E via $(x, y) \mapsto (N^2x, N^3y)$. We use Δ for the discriminant of E, i.e.

$$\Delta = -4A^3 - 27B^2.$$

Under this assumption we then have

Lemma 2.4. *Let $\mathfrak{p}$ be a prime ideal of K not dividing the discriminant Δ of E. Then L is not ramified at $\mathfrak{p}$.*

Proof. For $P \in E(K)$ let $M = K(Q \in E \,|\, mQ = P)$. It suffices to show that M is unramified at $\mathfrak{p}$ (since L is is the compositum of all such M).

Indeed let $D_\mathfrak{p}$ be the decomposition group of $\mathfrak{p}$ i.e. the subgroup of all $\sigma \in G$ leaving invariant one prime ideal (and hence all prime ideals) $\mathfrak{P}$ of M

above $\mathfrak{p}$. Let $I_\mathfrak{p}$ be the inertia group at $\mathfrak{p}$, i.e. the subgroup of $\sigma \in D_\mathfrak{p}$ such that $x^\sigma \equiv x \bmod \mathfrak{P}$ for all $x \in O$, where O is the ring of integers of M. That M is not ramified at $\mathfrak{p}$ is equivalent to the statement that $I_\mathfrak{p}$ is trivial.

For proving this we consider, $\widetilde{E}$, the curve obtained from E by reducing modulo $\mathfrak{P}$. More precisely we consider the following: If $P = [x : y : z]$ is a point of $E(M)$, then we may assume that x, y, z are in O, and at least one homogeneous coordinate is not divisible by $\mathfrak{P}$ (indeed take any homogeneous coordinates of P in M and divide by the the one with smallest $\mathfrak{P}$-order; since the new homogeneous coordinates are $\mathfrak{P}$-integral, we can find an integer $N \neq 0$ and not divisible by $\mathfrak{P}$ such that multiplication by N yields homogeneous coordinates in O). We then set $\rho(P) := [\widetilde{x} : \widetilde{y} : \widetilde{z}]$, where the tilde denotes the class modulo $\mathfrak{P}$. This does not depend on the special choice of homogeneous coordinates. The association $P \mapsto \tilde{P}$ thus defines a map

$$E(L_P) \to \widetilde{E}(O/\mathfrak{P}) = \{[\widetilde{x} : \widetilde{y} : \widetilde{z}] \,|\, y^2 z \equiv x^3 + Axz^2 + z^3 \bmod \mathfrak{P}\}.$$

It is easy to see that $E(O/\mathfrak{P})$ is a group (defined analogous to the group structure on $E(K)$), and that the reduction map is a group homomorphism. Moreover, it is a fundamental fact that the restriction of the reduction map to

$$E[m] \to \widetilde{E}(O/\mathfrak{P})$$

is injective if the discriminant of E is not divisible by $\mathfrak{P}$ (or, equivalently, not divisible by $\mathfrak{p}$). This is obvious for $m = 2$ (the case, which actually suffices to deduce the Mordell-Weil theorem). In this case $[0 : 1 : 0]$ and $[\alpha_i : 0 : 1]$ $(i = 1, 2, 3)$, with α_i denoting the roots of $f(x) := x^3 + Ax + B = 0$, are the points of $E[2]$ (recall that $\alpha_i \in K$ since $E[2] \subset E(K)$). Obviously they are in fact incongruent modulo $\mathfrak{P}$ if and only if $\mathfrak{P}$ does not divide the discriminant $\Delta = \prod_{i \neq j} (\alpha_i - \alpha_j)^2$. For general m see e.g. [Sil1], VII Proposition 3.1(b).

Let now $\sigma \in I_\mathfrak{p}$. Then

$$\rho(Q^\sigma - Q) = \rho(Q^\sigma) - \rho(Q) = 0$$

for all $Q \in E(M)$. On the other hand side, $Q^\sigma - Q \in E[m]$ for $mQ = P$. By the injectivity of the last map hence $Q^\sigma - Q = 0$. Thus σ is the identity on M, showing that $I_\mathfrak{p}$ is trivial and thus proving the theorem. □

Our information about L obtained so far suffices to prove that is is finite over K. One has the following general theorem:

Theorem 2.16. *Let L be an abelien extension of K with exponent m, and which is ramified only at a finite number of primes. Then L is a finite extension of K.*

Proof. Let S be the set of prime ideals of K, where L is ramified. By enlarging S we can assume that all prime ideals dividing m are contained in S. Moreover, by again enlarging if necessary, we can even more assume that the ring R of S-integers in K is a principal ideal domain. Indeed, let h be the class number of K, pick prime ideals $\mathfrak{p}_j$ $(1 \leq j \leq h)$ which represent the ideal classes of the class group of K, and adjoin to S all prime ideals conjugate to one of these prime ideals; clearly, $\mathfrak{p}_j^n R = R$ for all integers n (if $p \in \mathfrak{p}_j$ is a rational prime then $p^{-1} \in R$). But then, if $M \subseteq R$ is an ideal of R (and hence $M \cap O$ is an ideal of the ring of integers O of K), then, on writing $M \cap O$ as $M \cap O = \alpha \prod_j \mathfrak{p}_j^{n_j}$ with suitable integers n_j and suitable $\alpha \in O$, shows $(M \cap O)R = \alpha R$. But $(M \cap O)R = M$ (since, for each $\alpha \in M$, we can find a rational integer $N \in R$, only divisible by prime ideals in S, such that $N\alpha \in O$; but then $\alpha \in (M \cap O)R$ since $1/n \in R$).

Finally, we leave it to the reader to verify that, by adjoining mth roots of unity to K and L, one can assume without loss of generality that K contains all mth roots of unity. Or else the reader can restrict to the case of $m = 2$, where this is automatically satisfied, and which suffices for the proof Mordell-Weil theorem (and which in turn implies the weak version for arbitrary $m \geq 2$ anyway).

The main theorem of Kummer theory states that L is then a subfield of $K(\sqrt[m]{a} \,|\, a \in K)$ ([Lan2], VIII, §8) or any other reasonable text book including sections on Galois theory). Again, it is a straight-forward exercise in Galois theory to verify this statement for $m = 2$.

To begin with the proper proof of the desired theorem, we remark first of all that, for $a \in K$, the field $K(\sqrt[m]{a})$ is unramified at a prime $\mathfrak{p} \not| m$ if and only if $m | \operatorname{ord}_{\mathfrak{p}}(a)$ (Exercise).

Thus, if we let T be the set of classes $a(K^*)^m$ in $K^*/(K^*)^m$ such that $m | \operatorname{ord}_{\mathfrak{p}}(a)$ for all $\mathfrak{p} \notin S$, then

$$L \subseteq K(\sqrt[m]{a} \,|\, (aK^*)^m \in T).$$

We thus want have to show that T is finite. For this let R^* be the group of units of R. Clearly, $\operatorname{ord}_{\mathfrak{p}}(a) = 0$ for all $\mathfrak{p} \notin S$. Hence we have the natural map

$$R^* \to T.$$

We claim that it is surjective (for our special choice of R). Indeed, let $a \in K^*$ represent an element of T. Then the (fractional) R-ideal aR is the m th power of an R-ideal (consider the O-prime ideal decomposition of aO, multiply by R, and use that $\mathfrak{p}R = R$ for any $\mathfrak{p} \in S$). But R is a principal ideal domain, and hence $aR = b^m R$ for some $b \in K$, whence $a = b^m e$ for some unit $e \in R^*$, proving the surjectivity of our map. This map factorizes then to a surjective map $R^*/(R^*)^m \to T$.

By Dirichlet's S-unit theorem R^* is finitely generated (see [Lan3] V§1), hence $R^*/(R^*)^m$, and thus T too, is finite. □

2.8 Supplements

The Kummer pairing $E(K)/mE(K) \times G \to E[m]$ can be interpreted as injection

$$\delta_E : E(K)/mE(K) \to \mathrm{H}^1(G, E[m]).$$

Here $\mathrm{H}^1(G, E[m])$ is the first cohomology group of $G := \mathrm{Gal}(\overline{\mathbb{Q}}/K)$ acting on $E[m]$. Recall that, for any abelian group M which is a G-right module, this is the group

$$\mathrm{H}^1(G, M) = \frac{\{c : G \to M \,|\, c(\sigma\tau) = c(\sigma)^\tau + c(\tau)\}}{\{c : G \to M \,|\, \exists m \in M \,\forall \sigma \in G \,: c(\sigma) = m^\sigma - m\}}.$$

If $E[m] \subset E(K)$, as we assumed, then $\mathrm{H}^1(G, E[m])$ is nothing else than the group of homomorphisms $G \to E[m]$. Moreover, the map δ_E is nothing else as the map induced by the first connecting homomorphism, usually denoted δ, in the long exact sequence of homology groups

$$0 \to E[m](K) \to E(K) @ > \times m >> E(K) @ > \delta >> \mathrm{H}^1(G, E[m])$$

associated to the short exact sequence of G-modules

$$0 \to E[m] \to E @ > \times m >> E \to 0.$$

Note that the map δ_E exists for arbitrary E defined over K, not just for those with $E[m] \subset E(K)$. Along these lines the given proof of the Mordell theorem may be reinterpreted and reanalyzed in terms of Galois cohomology.

The approach to the weak Mordell theorem in section 2.3 using the map $E(\mathbb{Q})/2E(\mathbb{Q}) \to (\mathbb{Q}^*/\mathbb{Q}^*)^2$, can easily be generalized to arbitrary number fields (see [Lan1], V, §1), and it can also be generalized to arbitrary m (see e.g. [Sil1], X, Theorem 1.1). It is related to the second proof as follows.

By Hilbert's theorem 90 (which states $\mathrm{H}^1(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \overline{Q}^*) = 0$) we know that any homomorphism

$$\alpha : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \{\pm 1\}$$

is of the form $\alpha(\sigma) = \sqrt{a}^\sigma/\sqrt{a}$ with a suitable $a \in \mathbb{Q}^*$. Hence we have an isomorphism

$$\delta_K : \mathbb{Q}^*/\mathbb{Q}^{*2} \to \mathrm{Hom}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), \{\pm 1\}).$$

Suppose, we have a perfect pairing $e_2 : E[2] \times E[2] \to \{\pm 1\}$. Then we can define a unique map ν such that the following diagram is commutative:

$$\begin{tikzcd}
E(\mathbb{Q})/2E(\mathbb{Q}) \times E[2] \arrow[r, "\nu"] \arrow[d, "\delta_R \times 1"'] & \mathbb{Q}^*/\mathbb{Q}^{*2} \arrow[d, "\delta_K"'] \\
\mathrm{Hom}(G, E[2]) \times E[2] \arrow[r, "e_2'"'] & \mathrm{Hom}(G, \{\pm 1\})
\end{tikzcd}$$

Here $(G = \mathrm{Gal}(\overline{Q}/\mathbb{Q}),\ )$ and e_2' is the map induced by e_2, i.e. $e_2'(c, Q)(P) = e_2(c(P), Q)$ for all $P, Q \in E[2]$. Choosing a basis P_1, P_2 for $E[2]$, we then obtain an injection

$$\gamma : E(\mathbb{Q})/2E(\mathbb{Q}) \to (\mathbb{Q}^*/\mathbb{Q}^{*2})^2, \quad P \mapsto \big(\nu(P, P_1), \nu(P, P_2)\big).$$

Now, for e_2 one may take the so-called Weil pairing, which is defined as

$$e_2(P, Q) = g_Q(X + S)/g_Q(X),$$

where $g_Q \in K(E)$ is any function with divisor

$$\mathrm{div}(g) = \sum_{2R=Q} (R) - 4(O),$$

and where X is any point of E such that $g_Q(X + S)$ and $g_Q(X)$ are both different from 0 and ∞ (see any text book on (algebraic) elliptic curves) If $Q = (\alpha, 0)$ in affine coordinates, then it is not hard to check that $g_Q^2(V) = x(2V) - \alpha$ for all $V \in E$ (after suitably normalizing g_Q). Using this one can finally verify that γ is the map used in section 2.3.

2.9 Local decomposition

As in the case of algebraic numbers the canonical height on an elliptic curve has a decomposition into local contributions. In this section we describe the corresponding formulas. Again, we assume throughout that E is given by an equation of the form

$$E : y^2 = x^3 + Ax + B \qquad (A, B \in K),$$

where, as usual, K denotes a number field.

2.9.1 The Green's function of an elliptic curve

We start by describing the archimedian contributions. It is a well-known and classical fact that there exist a lattice in $\mathbb{C}$ of the form $L = \mathbb{Z}\tau + \mathbb{Z}$ with $\mathrm{Im}(\tau) > 0$ and a complex number $\lambda \neq 0$ such that the map

$$z \mapsto \begin{cases} [\wp(\tau, z) : \frac{1}{2}\wp'(\tau, z) : 1] & \text{if } z \notin L; \\ [0 : 1 : 0] & \text{if } z \in L \end{cases}$$

defines a surjective group homomorphism

$$\exp : \mathbb{C} \to E'(\mathbb{C})$$

with kernel L, where E' is the elliptic curve

$$E' : y^2 = x^3 + \lambda^4 Ax + \lambda^6 B.$$

Here $\wp(\tau, z)$, for fixed τ, as function of z, is the classical Weierstrass $\wp$ function associated to the lattice L, and $\wp'(\tau, z)$ is its derivative with respect to z. Thus, $\wp(\tau, z)$ is meromorphic in $\mathbb{C}$ with poles only in L, periodic with respect to L, and

$$\wp(\tau, z) = \frac{1}{z^2} + O(z) \qquad (z \to 0).$$

These three properties uniquely determine $\wp(\tau, z)$ (since the difference of any two such functions would be holomorphic on all of $\mathbb{C}$, periodic under L, hence bounded on $\mathbb{C}$, hence constant by the maximum principle, and finally equal to 0 because its Taylor development at $z = 0$ starts with positive powers of z). We use here the name exp because this is natural when viewing $E(\mathbb{C})$ as Lie group. Note that *exp* is continuous, when we equip $E'(\mathbb{C})$ with the natural topology (inherited from the natural quotient topology of $\mathbb{P}^2(\mathbb{C}) = (\mathbb{C}^3 \setminus \{0\})/\mathbb{C}^*$). To check this at points in L write

$$[\wp(\tau, z) : \frac{1}{2}\wp'(\tau, z) : 1] = [\frac{\wp(\tau, z)}{\frac{1}{2}\wp'(\tau, z)} : 1 : \frac{1}{\frac{1}{2}\wp'(\tau, z)}],$$

an let z tend towards a point in L.

Clearly E' and E are isomorphic (as elliptic curves over $\mathbb{C}$) via the map $(x, y) \mapsto (\lambda^2 x, \lambda^3 y)$. For the following we assume that $E = E'$ (and hence $\lambda = 1$). Of course, then A, B are not necessarily algebraic numbers.

One can even more introduce a natural structure of Riemann surface on $\mathbb{C}/L$ and on $E(\mathbb{C})$ so that the map $\mathbb{C}/L \to E(\mathbb{C})$ becomes an isomorphism of Riemann surfaces. The map exp induces an isomorphism of fields

$$\exp^* : K(E)_{\mathbb{C}} \to \mathcal{M}(L),$$

where $K(E)_{\mathbb{C}}$ is the field if algebraic functions on E, considered as algebraic curve over $\mathbb{C}$, and where $\mathcal{M}(L)$ is the field of meromorphic functions on $\mathbb{C}$ which are periodic with respect to L.

We use $\sigma(\tau, z)$ for the Weierstrass σ function associated to L. It is uniquely characterized by the fact that, as function in z, it is odd and holomorphic on $\mathbb{C}$, satisfies $\sigma(\tau, z) = z + O(z^2)$ $(z \to 0)$, and its second logarithmic derivative equals $\wp(\tau, z)$. Setting

$$q = \mathrm{e}^{2\pi i\tau}, \quad \zeta = \mathrm{e}^{2\pi i z},$$

one has the following explicit formula ([Skor], Appendix 1)

$$\sigma(\tau, z) = \mathrm{e}^{-2\pi i \frac{\eta'}{\eta}(\tau) z^2} \frac{\zeta^{1/2} - \zeta^{-1/2}}{2\pi i} \prod_{n \geq 1} \frac{(1 - q^n\zeta)(1 - q^n\zeta^{-1})}{(1 - q^n)^2},$$

$$\eta(\tau) = q^{1/24} \prod_{n \geq 1} (1 - q^n).$$

(Here η' is the ordinary derivative of η with respect to τ.) It is straightforward that the right hand side of this formula satisfies in fact all the listed properties, which proves the existence of $\sigma(\tau, z)$ (and $\wp(\tau, z)$) and, by the uniqueness, the identity in question. We leave the details to the reader (or see [Skor], Appendix 1). We cite without proof the following lemma (see)

Lemma 2.5.

$$(2\pi i)^{12}\eta^{24}(\tau) = \mathrm{disc}(x^3 + Ax + B) = -(4A^3 + 27B^2).$$

Instead of in $\sigma(\tau, z)$, we are more interested in the so-called Siegel function

$$S(z) = q^{\frac{1}{12}}\zeta^{-\frac{1}{2}}(\zeta - 1) \prod_{n \geq 1} (1 - q^n\zeta)(1 - q^n\zeta^{-1}).$$

We suppress the dependence of τ. Note that $S(z)$, considered as function of z is nothing else but $\sigma(\tau, z)$, up to multiplication by trivial factors. The important point is that $S(z)$ has a nicer transformation law under L than $\sigma(\tau, z)$. Namely, one has

Lemma 2.6.

$$S(z + 1) = -S(z), \qquad S(z + \tau) = -q^{-\frac{1}{2}}\zeta^{-1}S(z).$$

Proof. This can be verified by a straight-forward calculation. □

From this we deduce that

$$G(z) := \mathrm{e}^{-\pi\frac{y^2}{v}}|S(z)|$$

is periodic with respect to L. Here y and v are the imaginary parts of z and τ, respectively.

factors through a function on $\mathbb{C}/L$. This function is Green's function associated to E. Its important property is

Theorem 2.17. *Let $f \in K(E)_{\mathbb{C}}$, let $D = \sum_{j=1}^{r} n_j(P_j)$ $(n_j \in \mathbb{Z},\ P_j \in E(\mathbb{C}))$ its divisor, and let $P_j = \exp(z_j)$ with suitable $z_j \in \mathbb{C}$. Then there exists a constant c such that*

$$|f(\exp(z))| = c\prod_{j=1}^{r} G(z-z_j)$$

for all $z \in \mathbb{C}$.

Proof. The function

$$g(z) := f(\exp(z))/\prod_{j=1}^{r} S(z-z_j)^{n_j}$$

is holomorphic on $\mathbb{C}$ and has no zeroes. From this it is easy to verify that

$$\tilde{g}(z) := \log g(z) + \pi\frac{1}{v}\sum_{j=1}^{r} n_j \operatorname{Im}(z-z_j)^2$$

is harmonic (though $G(z)$ itself, because of the factor $\mathrm{e}^{-\pi y/v}$, is not harmonic). Note that

$$\sum_{j=1}^{r} n_j \operatorname{Im}(z-z_j)^2$$

is harmonic since D, as divisor of a function on $K(E)_{\mathbb{C}}$, satisfies $\deg D = \sum_{j=1}^{r} n_j = 0$.

But $\tilde{g}$ is periodic with respect to L, hence bounded on $\mathbb{C}$, and thus constant by the maximum principle. □

As corollary we obtain

Corollary 2.17.1.

$$|\wp(z) - \wp(a)| = |\Delta|^{\frac{1}{6}}\frac{G(z-a)G(z+a)}{G(z)^2G(a)^2}.$$

Proof. By the foregoing theorem we have, for fixed a and all z

$$|\wp(z) - \wp(a)| = c\frac{G(z-a)G(z+a)}{G(z)^2G(a)^2}$$

with a suitable constant c. Now, if we multiply by $|z|^2$ and let z tend to 0, then the left hand side tends to 1. For the right hand side the limit is

$$c \cdot \lim_{z\to 0} \frac{|z|^2}{G(z)^2} = c/(2\pi|\eta|^2)^2,$$

which proves the lemma. □

We finally introduce the so-called Néron function on $E(\mathbb{C})\setminus\{0\}$ by setting

$$\lambda(P) := -\log G(z),$$

where $P = \exp(z)$ (this does not depend on a particular choice of z since $G(z)$ is periodic with respect to L.)

Theorem 2.18. *The Néron function satisfies the following three conditions:*

1. *λ is continuous and is bounded on the complement of every open neighbourhood of 0.*
2. *The limit $\lim_{P\to 0}\left(\lambda(P) + \frac{1}{2}\log|x(P)|\right)$ exists and is finite.*
3. *For all $P, Q \in E(\mathbb{C})$ such that $P, Q, P+Q, P-Q \neq 0$ one has*
$$\lambda(P+Q) + \lambda(P-Q) = 2\lambda(P) + 2\lambda(Q) - \log|x(P) - x(Q)| + \frac{1}{6}\log|\Delta|.$$

Moreover, λ is the only function on $E(\mathbb{C})\setminus\{0\}$ satisfying these conditions.

Proof. Property (i) to (iii) follow immediately from the corresponding properties for $-\log G(z)$ on setting $P = \exp(z)$ and $Q = \exp(a)$, so that, in particular $x(P) = \wp(\tau, z)$.

For proving the uniqueness statement we note that the difference f of any two functions satisfying the three properties can be continuously extended to 0 (by (ii)), is hence bounded on $E(\mathbb{C})$ (by (i)), and satisfies the parallelogram law

$$f(P+Q) + f(P-Q) = 2f(P) + 2f(Q)$$

(by (iii)), by continuity even for all P, Q. In particular, $f(0) = 0$ (set $P = Q = 0$), hence $f(2P) = 4f(P)$ (set $P = Q$), and then $f(2^nP) = 4^nf(P)$ for all P and n. Letting n tend to infinity and observing that $f(2^nP)$ remains bounded it follows $f(P) = 0$. □

If $E' : y^2 = x^3 + A'x + B'$ is an elliptic curve isomorphic to E, say via $\alpha : (x, y) \mapsto (a^2x, a^3y)$, we transfer λ to a function λ' on E' by setting $\lambda' = \lambda \circ \alpha$. Note that the conditions (i) and (iii) remain literally valid for the new function λ' on E'. Indeed, if we write $x'(P)$ for the first coordinate function on E', then we have $(x \circ \alpha)(P) = a^2x'(P)$, whereas the discriminant Δ' of E' is

$$\Delta' = -(4A'^3 + 27B'^2) = -a^{12}(4A^3 + 27B^2) = a^{12}\Delta$$

(since $A' = a^4A$ and $B' = a^6B$) Hence

$$\log|x(\alpha(P)) - x(\alpha(Q))| - \frac{1}{6}|\Delta| = \log|x'(P) - x'(Q)| - \frac{1}{6}|\Delta'|.$$

Hence we can summarise by saying that on each elliptic curve E defined over $\mathbb{C}$, given by a Weierstrass equation with discriminant Δ, there is a unique function $\lambda : E(\mathbb{C}) \setminus \{0\} \to \mathbb{R}$ which satisfies properties (i) to (iii).

The condition (iii) can be replaced by another one, which is technically simpler to verify.

Theorem 2.19. *Let $E : y^2 = x^3 + Ax + B$ an elliptic curve defined over $\mathbb{C}$. Then the Néron function λ is the unique function $\lambda : E(\mathbb{C}) \setminus \{0\} \to \mathbb{R}$ which satisfies conditions (i), (ii) of Theorem 2.18 and the condition:*

(iii)' For all $P \in E(\mathbb{C})$ such that $2P \neq 0$ one has

$$\lambda(2P) = 4\lambda(P) - \log|2y(P)| + \frac{1}{4}\log|\Delta|.$$

Proof. The proof that λ is uniquely determined by (i),(ii) and (iii)' is exactly the same as the uniqueness proof of the preceding theorem. In fact, all we used from (iii) is that the difference f of any two Néron functions satisfies $f(2P) = 4f(P)$, which is already implied by (iii)'.

For proving (iii)' we assume first of all as before that $E(\mathbb{C})$ is the homomorphic image under the exponential map exp with respect to a suitable lattice $L := \mathbb{Z}\tau + \mathbb{Z}$. Then, setting $P = \exp(z)$ (so that $2P = \exp(2z)$ and $\frac{1}{2}\wp'(\tau, z) = y(P)$) we have to prove

$$|\wp'(\tau, z)| = |\Delta|^{\frac{1}{4}}\frac{G(2z)}{G(z)^4}.$$

But this follows immediately from Theorem 2.17 on comparing divisors on both sides (note that $G(2z) = 0$ if and only if $z \in \frac{1}{2}L$), which proves the identity up to multiplication by a constant, and by multiplying by $|z|^3$ and letting z tend to 0.

Finally one proves as for condition (iii) that (iii)' remains literally valid for the Néron function on an arbitrary elliptic curve (over $\mathbb{C}$) in Weierstrass normal form. □

2.9.2 The Néron functions associated to places

In this section we return again to an elliptic curve E defined over a number field K, say

$$E : y^2 = x^3 + Ax + B, \quad \Delta = -(4A^3 + 27B^2), \qquad (A, B \in K)$$

If v is a place of (i.e. equivalence class of valuations on) K, then we use $\|.\|_v$ for that valuation in v, whose restriction to $\mathbb{Q}$ equals the ordinary p-adic valuation $|\cdot|_p$ for some prime number p or the usual archimedean absolute value on $\mathbb{Q}$. We then have, with a suitable integer $n_v \geq 1$, the identity

$$|\alpha|_v = \|\alpha\|_v^{n_v}$$

for all $\alpha \in K$. We use K_v for the v-adic completion of K, and we the same symbol for the extension of $\|\cdot\|_v$ to K_v.

Generalising the theorem of the last section one can prove:

Theorem 2.20. *Let v be a place of K. Then there exists a unique function $\lambda_v : E(K_v)\setminus\{0\} \to \mathbb{R}$ satisfying properties (i) to (iii) of Theorem 2.18 with $\mathbb{C}$ replaced by K_v and the complex absolute value replaced by $\|\cdot\|_v$. The function λ_v can also be characterised as the unique real-valued function on $E(K_v)\setminus\{0\}$ which satisfies conditions (i), (ii) and condition (iii)' (of Theorem 2.19 with the same replacements as before). Assume that A and B are integral. Then, for all but finitely many v one has*

$$\lambda_v(P) = \frac{1}{2}\max(0, \log \|x(P)\|_v)$$

for all $P \in E(K_v) \setminus \{0\}$.

The function λ_v is called the local Néron function on E associated to v. The uniqueness of λ_v follows literally as in the proof of Theorem 2.18. If L is an extension of K, and if w is a place of L over v, then, since the restriction of λ_w to $E(K_v) \setminus \{0\}$ satisfies (i) to (iii), whence $\lambda_w(P) = \lambda_v(P)$ for $P \in E(K) \setminus \{0\}$.

If v is archimedean, i.e. if $K_v = \mathbb{C}$ or $K_v = \mathbb{R}$, then the existence of λ_v is ensured by Theorem 2.18. We shall not give the complete proof of the preceding theorem in the case of a non-archimedean v, but refer to the literature (cf. [Sil2]).

Here we content ourselves to prove the following theorem, which implies a part the preceding one for non-archimedean v where E has good reduction (and a little bit more). To state this theorem we need some notation.

Let $v \in P_K$ non-archimedean and assume that A and B are v-integral (i.e. $\|A\|_v, \|B\|_v \leq 1$). Let

$$O_v = \{x \in K_v \mid \|x\|_v \leq 1\}, \quad \mathfrak{m}_v = \{x \in K_v \mid \|x\|_v < 1\}.$$

Denote by $\widetilde{E}$ the curve over the field $O_v/\mathfrak{m}_v$ obtained from E by reducing A and B modulo the maximal ideal $\mathfrak{m}_v$. We have the map (in fact a homomorphism)

$$E \to \widetilde{E}, \qquad P \mapsto \widetilde{P}$$

obtained by reducing modulo $\mathfrak{m}_v$ (as explained in the proof of Lemma 2.4). We set

$$E_0(K_v) = \{P \in E(K_v) \mid \widetilde{P} \text{ is a nonsingular point of } \widetilde{E}\}.$$

It can be proved that this is a subgroup of $E(K_v)$ (see e.g. [Sil1], VII §2).

Theorem 2.21. *Let $v \in P_K$ non-archimedean, and assume that A and B are v-integral. Then*

$$\lambda_v(P) = \frac{1}{2}\max(\log \|x(P)\|_v, 0) - \frac{1}{12}\log \|\Delta\|_v$$

for all $P \in E_0(K_v) \setminus \{0\}$.

Proof. Denote the function on $E(K_v) \setminus \{0\}$ defined by the right hand side of the desired formula by λ. Clearly, λ satisfies properties (i) and (ii) of the local Néron function. Writing $|x|$ for $\|x\|_v$ we shall show the duplication formula

$$\lambda(2P) = 4\lambda(P) - \log|2y(P)| + \frac{1}{4}\log|\Delta|$$

for all $P \in E_0(K_v) \setminus \{0\}$.

This than implies that the restriction of λ_v to $E_0(K_v) \setminus \{0\}$ equals λ by the usual argument. Indeed, the difference $f = \lambda - \lambda_v$ extends to a continuous and bounded function on all of the subgroup $E_0(K_v)$ of $E(K_v)$. One has $f(2P) = 4f(P)$, by continuity even if $P = 0$ or $2P = 0$. But then $f(P) = 4^{-n}f(2^nP)$ for all n, which implies $f(P) = 0$ since f is bounded.

To prove the duplication formula for λ we note first of all (writing $x_1 = x(P)$ and $y_1 = y(P)$) that

$$x(2P) = -2x_1 + \frac{F_x(P)^2}{F_y(P)^2} = \frac{x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2}{4y_1^2} =: \frac{\phi}{\psi},$$

where $F(x,y) = y^2 - (x^3 + Ax + B)$ and F_x, F_y denote partial derivatives.

Hence, the duplication formula is equivalent to

$$\frac{1}{2}\max(\log|\phi| - \log|\psi|, 0) = 2\max(\log|x_1|, 0) - \log|2y_1|,$$

which, using $|\psi|^2 = |2y|$, can be written as

$$\max(|\phi|, |\psi|) = \max(|x_1|^4, 1)$$

Assume, first of all that $|x_1| > 1$. Then, using that A, B are v-integral, we have $|\phi| = |x_1|^4$ and $|\psi| = |4y_1^2| = |4(x_1^3 + Ax_1 + B)| = |4x_1^3| < |x_1|^4 = |\phi|$. Hence the desired identity is true.

Now assume that x_1 is v-integral. Since A, B are v-integral y_1 is then v-integral too, in particular, we have $\widetilde{P} = [x_1 + \mathfrak{m}_v : y_1 + \mathfrak{m}_v : 1]$. We shall now use that $\widetilde{P}$ is a non-singular point of the reduced curve $\widetilde{E}$. This is equivalent to $|F_x(x_1, y_1)| = 1$ or $|F_y(x_1, y_1)| = 1$. Since

$$\phi = F_x(P)^2 - 2x_1F_y(P)^2 \quad \psi = F_y(P)^2$$

this implies that indeed $\max(|\phi|, |\psi|) = 1$. □

Note that in the case of good reduction, i.e. if $\|\Delta\|_v = 1$, we have the explicit formula

$$\lambda_v(P) = \frac{1}{2}\max(\log\|x(P)\|_v, 0),$$

and that we have actually proved that the right hand side satisfies the defining conditions (i), (ii) and (iii)' of the local Néron function at v.

2.9.3 The decomposition formula

Using the local Néron functions λ_v we can finally give the desired local decomposition of the canonical height h.

Theorem 2.22. *Let E be an elliptic curve defined over the number field K, let h be the canonical height on E, and, for each $v \in P_K$ let λ_v be the local Néron height function associated to v. Then*

$$h(P) = \frac{1}{[K:\mathbb{Q}]}\sum_{v\in P_K}\lambda_v(P)$$

for all $P \in E(K) \setminus \{0\}$.

Proof. Note that by Theorem 2.20 for each $P \in E(K)$, $P \neq 0$ we have $\lambda_v(P) = \frac{1}{2}\max(\log \|x(P)\|_v, 0)$ for almost all v. Hence the sum on the right hand side of the desired formula is actually finite (and hence well-defined). Denote by $h'(P)$ the function on $E(K)$ defined by the right hand side of the desired formula if $P \neq 0$, and such that $h'(0) = 0$.

To prove $h = h'$ it suffices to prove that $|h'(P) - \frac{1}{2}h_x(P)|$ is bounded and that $h'(2P) = 4h'(P)$ (see Theorem 2.13); here the bars denote the ordinary absolute value on $\mathbb{R}$.

The latter follows, for $2P \neq 0$, immediately from

$$\lambda_v(2P) = 4\lambda_v(P) - \log \|2y(P)\|_v + \frac{1}{4}\log \|\Delta\|_v$$

and the product formula (here written additively)

$$\sum_{v \in P_K} n_v \log \|x\|_v = 0,$$

valid for all $x \in K$, $x \neq 0$. For $P = 0$ its is trivially true since $h'(0) = 0$ by definition. For $2P = 0$ and $P \neq 0$ we have to show $h'(P) = 0$. This can be done e.g. by the triplication formula $\lambda_v(3P) = \lambda_v(P) + \log \|f(P)\| + \frac{2}{3}\log \|\Delta\|$ valid for all P with $3P \neq 0$ (cf. [Sil2], Exercise 6.4 (e); here $f \in K(E)$ independent of P).

From property (i) and (ii) of the Néron function we deduce the existence of constants c_v such that

$$-c_v \leq \lambda_v(P) - \frac{1}{2}\log\max(\|x(P\|_v, 1) \leq c_v$$

for all $v \in P_K$ and all $P \in E(K) \setminus \{0\}$. Even more, by the last theorem we can and will choose $c_v = 0$ for all but a finite number of v. Multiplying by $n_v/[K : \mathbb{Q}]$ and summing over all v then yields

$$|h'(P) - \frac{1}{2}h_x(P)| \leq \frac{1}{[K : \mathbb{Q}]}\sum_{v \in P_K} n_v c_v,$$

and hence the desired inequality. □

Part 3

Appendix: Exercises

The following exercises were given to the student at the end of the course as a written examination (in French). However, they supplement some of the threads of these notes and may hence be of independent interest.

3.1 Mesure de Mahler de polynômes en plusieurs variables

Pour un polynôme $P \in \mathbb{C}[X_1, \ldots, X_n]$, $P \neq 0$, on pose

$$\mu(P) := \exp\Big(\int_0^1 \cdots \int_0^1 \log |P(\mathrm{e}^{2\pi i t_1}, \ldots, \mathrm{e}^{2\pi i t_n})| \, dt_1 \cdots dt_n\Big),$$

et on pose $\mu(0) = 0$. Dans l'exercice suivant la formule du cours

$$\int_0^1 \log |\alpha - \mathrm{e}^{2\pi i t}| \, dt = \log_+ |\alpha|$$

sera utile[1].

(i) En utilisant que $\mu(f) \geq |a_d|$ pour tout polynôme $f(x) = a_d X^d + \cdots + a_0$ en une variable, montrer par récurrence sur n que $\mu(P) \geq 1$ si P a des coefficients entiers.

(ii) Montrer : Si $|\alpha_k| \geq \sum_{j=0,\, j\neq k}^{n} |\alpha_j|$ pour un $0 \leq k \leq n$, alors

$$\mu(a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = |a_k|.$$

En déduire $\mu(X_1 + X_2 + k) = |k|$ pour $|k| \geq 2$.

[1] Nous utilisons la notation $\log_+ x = \log \max(x, 1)$ $(x \in \mathbb{R},\ x > 0)$.

(iii) Calculer $\mu(X_1 + X_2)$.

(iv) Montrer d'abord que $\mu(X_1+X_2+1) = \int_{-1/3}^{1/3} \log|1+\mathrm{e}^{2\pi it}|\,dt$. Devélopper $\log|1 + \mathrm{e}^{2\pi it}| = \mathrm{Re}\log(1 + \mathrm{e}^{2\pi it})$ comme série en puissance de $\mathrm{e}^{2\pi it}$, échanger l'intégration et sommation (on admets la justification), et en déduire que

$$\log\mu(X_1 + X_2 + 1) = \frac{3\sqrt{3}}{4\pi} L(2, \left(\frac{\cdot}{3}\right)),$$

$$\text{où} \qquad L(s, \left(\frac{\cdot}{3}\right)) := \sum_{n=1}^{\infty} \left(\frac{n}{3}\right) n^{-s} \quad (s > 1).$$

On utilisera $\sum_n \left(\frac{n}{3}\right) n^{-s} = -4^{-s} \sum_n \left(\frac{n}{3}\right) n^{-s} + \sum_{n \text{ impair}} \left(\frac{n}{3}\right) n^{-s}$.

3.2 Calcul rapide de l'hauteur canonique

Soit $E : y^2 = x^3 + Ax + B$ une courbe elliptique définie sur $\mathbb{Q}$. Dans cet exercice on se propose de démontrer une formule pour l'hauteur canonique h sur E, qui peut être utile pour un calcul rapide. Pour simplifier nous supposons le suivant :

1. $A, B \in \mathbb{Z}$.
2. On a $f(x) := x^3 + Ax + B = (x - \alpha)(x - \overline{\alpha})(x - \beta)$ avec $\alpha \notin \mathbb{R}$ et $\beta > 0$.

Soit $\phi(x)$ le polynôme (de degré 4) tel que

$$x(2P) = \frac{\phi(x(P))}{4f(x(P))}$$

pour tout $P \in E(\mathbb{R})$, $P \neq 0$. Nous posons $h'(0) = 0$, et pour $P \in E(\mathbb{Q})$, $P \neq 0$, $x(P) = \frac{a}{b}$ avec $a, b \in \mathbb{Z}$ tels que $\mathrm{pgcd}(a, b) = 1$ nous posons

$$h'(P) = \log|a| + \sum_{n=0}^{\infty} \frac{1}{4^{n+1}} \log|\phi(x_n)/x_n^4|$$

$$\text{où} \quad x_0 = x(P), \quad x_{n+1} = \frac{\phi(x_n)}{4f(x_n)} \ (n \geq 0).$$

(i) Montrer que, pour $x \in \mathbb{R}$, $x > \beta$, on a $f(x) \neq 0$ et $\phi(x)/4f(x) \geq \beta$. Calculer $\phi(x)$ et montrer que $\phi(x)/x^4 \to 1$ pour $t \to \infty$ et $\phi(\beta)/\beta^4 > 0$. En déduire qu'il existe des constantes $c_1 > 0$ et c_2 telles que l'on a $c_1 \leq \phi(x)/x^4 \leq c_2$ pour tout $x \geq \beta$.

(ii) Déduire de (i) que la somme qui définit $h'(P)$ est bien-définie et converge absolument (en fait très rapidement).

(iii) En utilisant sans preuve le fait que $\mathrm{pgcd}(\phi(a/b)b^4, 4f(a/b)b^4) = 1$, montrer que $h'(2P) = 4h'(P)$.

(iv) Montrer : Il existe une constante c tel que $|h'(P) - \log\max(|a|, |b|)| \leq c$. (Ici l'estimation de (i) sera encore utile).

(v) Déduire de (iii) et (iv) que $h(P) = \frac{1}{2}h'(P)$.

3.3 Fonctions de Néron

Soit $E : y^2 = x^3 + Ax + B$ une courbe elliptique avec discriminant Δ définie sur le corps de nombre K_0, soit $|\cdot|$ une valuation de K_0 et K la complétion de K_0 par rapport à $|\cdot|$. Nous allons montrer dans cet exercice l'existence de la fonction de Néron en $|\cdot|$. Plus précisemment, nous nous proposons de montrer qu'il existe une fonction $\lambda : E(K) \setminus \{0\} \to \mathbb{R}$ tel que

1. λ est continu est borné sur le complément de tout voisinage de 0.
2. $\lim_{P\to 0}(\lambda(P) - \frac{1}{2}\log|x(P)|)$ existe (et est fini).
3. $\lambda(2P) = 4\lambda(P) - \log|2y(P)| + \frac{1}{4}\log|\Delta|$ pour tout $P \in E(K)$ tel que $2P \neq 0$.

(0) Montrer que $x(2P) = \frac{\phi(x(P))}{4f(x(P))}$ pour tout $P \in E$, où $f(x) = x^3 + Ax + B$ et $\phi(x) = -8xf(x) + f'(x)^2$.

(i) Pour $P \in E(K)$, $2P \neq 0$ on pose

$$f(P) := \frac{1}{2}\log_+|x(2P)| - 2\log_+|x(P)| + \log|2y(P)| - \frac{1}{4}\log|\Delta|.$$

Montrer que $g(P) := \exp(f(P))$ peut être prolongé à une fonction continue sur $E(K)$. Calculer $g(0)$ et en déduire qu'il existe un $c > 0$ tel que $g(P) > 0$ pour $|x(P)| > c$.

(ii) Montrer que les polynômes $\phi(x)$ et $4f(x)$ sont relativement premiers, et qu'ils existent donc des polynômes $a(x)$, $b(x)$ tel que $1 = a\phi + 4bf$, En déduire que $g(P) > 0$ pour $x(P) \leq c$ (avec le c de (i)).

(iii) Déduire de (i) et (ii) que $f(P)$ peut être prolongé uniquement à une fonction continue et bornée sur tout $E(K)$.

(iv) Montrer, en utilisant (iii), que la somme

$$\mu(P) := \sum_{n=0}^{\infty} 4^{-n} f(2^n P)$$

converge pour tout $P \in E(K)$ et définit une fonction continue et bornée $\mu : E(K) \to \mathbb{R}$ qui satisfait $f(P) = 4\mu(P) - \mu(2P)$ pour tout $P \in E(K)$.

(v) Montrer, en résumant, que la fonction $\lambda(P) := \lambda_1(P) + \mu(P)$, définie pour $P \in E(K)$, $P \neq 0$, satisfait aux propriétés 1. à 3.

Bibliography

[Ahlf] Lars V. Ahlfors, Complex Analysis. 2nd edition, McGraw-Hill Kogakusha, Tokyo 1966. 4

[BeZa] F. Beukers and D. Zagier, Lower bounds of heights of points on hypersurfaces, Acta Arith. LXXIX (1997), 103–111. 29

[Bomb] E. Bombieri, The Mordell conjecture revisited, preprint *???* 39

[Boyd] D.W. Boyd, reciprocal numbers having small measure I, II, Comp. Math. 35 (1980), 1361–1377 and 53 (1989), 355–357, S1–S6. 8

[CoDi] H. Cohen and F. Diaz y Diaz, A polynomial reduction algorithm Sém. Théor. Nombres Bordeaux (Sér. 2) 3 (1991), 351-360. 6

[Dobr] E. Dobrowolski, On a question of Lehmer and the number of irreducible factors of a polynomial, Acta Arith. XXXIV (1979). 28

[Doch] C. Doche, Thèse de troisième cycle en préparation, Bordeaux 1998. 29

[FeTo] M. Fekete and G. Szegö, On algebraic equations with integral coefficients whose roots belong to a given point set, Math. Zeitschr. 63 (1955), 158–172. 19, 21

[Heck] E. Hecke, Vorlesungen über die Theorie der algebraischen Zahlen, Chelsea, New York 1970. 12

[HoSk] G. Hoehn and N-P. Skoruppa, Un résultat de Schinzel, Journ. Théor. Nombres Bordeaux 5 (1993), 185. 14

[Lan1] S. Lang, Fundamentals of Diophantine Geometry, Springer, New York 1983. 41, 52

[Lan2] S. Lang, Algebra, Addison-Wesley, Reading 1978. 51

[Lan3] S. Lang, Algebraic numbers, Addison Wesley, Reading 1964. 52

[Lvin] M. Langevin, *???* (ref. to Langevin's theorem) 18

[Lehm] D.H. Lehmer, Factorization of certain cyclotomic functions, Ann. Math. 34 (1933), 461–479. 7

[Loub] R. Louboutin, Sur la mesure de Mahler d'un nombre algébrique, C.R.Acad.Sci. Paris 296 (1983), 707–708. 28

[Mani] *???* (ref. to "For any K and any prime number p there exists a constant N such that the p-part of $E(K)_{\mathrm{tor}}$, for any E/K, is bounded to above by N.") 45

[Mazu] *???* (ref. to the classification of $E(\mathbb{Q})_{\mathrm{tor}}$.)

[Neuk] J. Neukirch, Algebraische Zahlentheorie. Springer, Berlin 1992. 2

[Pari] BaBeBeCoOl, Computer algebra package for number theorists, Bordeaux 1989-1998. 6

[Schi] A. Schinzel, On the product of the conjugates outside the unit circle of an algebraic number, Acta Arithmetica 24 (1973), 385–399. 14, 29

[Sieg] C.L. Siegel, Algebraic integers whose conjugates lie in the unit circle, Duke M. J. 11 (1944), 597–602 or No. 46 in gesammelte Abhandlungen. 9, 28

[Sil1] J.H. Silverman, The Arithmetic of Elliptic Curves, Springer, New-York 1986. 50, 52, 60

[Sil2] J.H. Silverman, Advanved Topics in the Arithmetic of Elliptic Curves, Spinger, New-York 1994. 59, 62

[Skor] N-P. Skoruppa, Modular forms *in* Hirzebruch, Berger and Jung, Manifolds and Modular Forms, Vieweg, Braunschweig 1992. 55

[Smy1] C.J. Smyth, On the product of the conjugates outside the unit circle of an algebraic integer, Bull. London Math. Soc. 3 (1971), 169–175. 23

[Smy2] C.J. Smyth, On the Mahler measure of the composition of two polynomials, Acta Arith. LXXIX (1997), 239–247. 29

[Weil] A, Weil, *???* (ref. to Weil's study of "divisor → line bundle → projective embedding → height".) 41

[Zagi] D. Zagier, Algebraic numbers close to both 0 and 1, Math. Comp. 61 (1993), 485–491. 14, 15, 31

[Zhan] S. Zhang, Positive line bundles on arithmetic surfaces, preprint, Princeton 1992. *???* 15