ANALYTIC AND COMBINATORIAL NUMBER THEORY II

(Lecture Notes)

MARTIN KLAZAR

These are lecture notes for the summer semester 2010 of the course *Analytic and combinatorial number theory* (NDMI045, *Analytická a kombinatorická teorie čísel*) which I have been teaching on the Faculty of Mathematics and Physics of the Charles University in Prague. In the second booklet (the first one, [22], was for summer semester 2008) we learn the theorems due to Thue (finiteness of solution set of Thue equation), Dirichlet (infinitude of primes in arithmetic progression) and Gel'fond and Schneider (transcendence of α^β for algebraic α and β).

July 2010 Martin Klazar

Contents

Notation

(a,b)	the greatest common divisor or an ordered pair
$a \mid b$	a divides b
$\mathrm{con}(\alpha)$	conjugates of α, p. 34
$\mathbb{C}$	complex numbers
χ	characters of finite abelian groups, p. 18
$\deg(\cdot)$	the degree of a polynomial
$\exp(z)$	$\sum_{n\geq 0} z^n/n!$
$\varphi(m)$	the Euler function, p. 18
G^*	the group of characters of G, p. 18
$G(K)$	embeddings of K in $\mathbb{C}$, p. 34
$h(\alpha)$	the size of α, p. 36
K_I	K-integers, p. 36
$[L:K]$	the degree of L over K, p. 33
$\Lambda(n)$	von Mangoldt function, p. 23
$\mu(m)$	Möbius function, p. 22
$\mathbb{N}$	$\{1,2,3,\dots\}$
$\mathbb{N}_0$	$\{0,1,2,\dots\}$
$\mathrm{ord}_p(a/b)$	the order of p in a/b, p. 27
p,q	in Chapter 2 denote prime numbers
$\lVert P\rVert$	the norm of a polynomial, p. 4
$\mathbb{Q}$	the set of rational numbers
$\mathbb{R}$	the set of real numbers
$\lvert X\rvert$, $\#X$	the cardinality of a set or sequence
$\mathbb{Z}$	the integers, $\{\dots,-2,-1,0,1,2,\dots\}$

Chapter 1

Thue's theorem on Diophantine equations

OM

EN GENEREL I STORE HELE TAL

ULØSBAR LIGNING

Theorem.

Er $F(x)$ en hvilkensomhelst hel irreduktibel funktion i x med hele koefficienter og af r'te grad, hvor $r > 2$, da har ligningen

$$q^r F(\tfrac{p}{q}) = c \qquad \cdots\cdots\cdots\cdot \ (1)$$

hvor c er et vilkaarlig opgivet helt tal, kun et begrændset antal løsninger i hele tal p og q.

beginning of the article [40] of A. Thue, by [43, pp. 219–231]

Über Annäherungswerte algebraischer Zahlen.

Von Herrn *Axel Thue* in Kristiania.

Theorem I. Bedeutet ρ eine positive Wurzel einer ganzen Funktion vom Grade r mit ganzen Koeffizienten, so hat die Relation

$$0 < |q\rho - p| < \frac{c}{q^{\frac{r}{2}+k}}, \qquad (1.)$$

wo c und k zwei beliebig gegebene positive Größen bezeichnen, nich unendlich viele Auflösungen in ganzen positiven Zahlen p und q.

(...)

Theorem IV. *Die Gleichung*

$$U(p, q) = c,$$

wo c eine gegebene Konstante ist, während U eine in bezug auf p und q ganze homogene und irreduktible Funktion mit ganzen Koeffizienten bedeutet, besitzt nicht unendlich viele Auflösungen in ganzen positiven Zahlen p und q, wenn der Grad von U größer als 2 ist.

parts of the article [41] of A. Thue, by [43, pp. 232–253]

In 1908, Norwegian mathematician Axel Thue (1863–1922) proved in [40] a deep result on finiteness of solution sets of a large class of Diophantine equations, nowadays called after him *Thue equations.*

Theorem 1.0.1 (Thue, 1908) *The Diophantine equation*

$$P(x, y) = a_d x^d + a_{d-1} x^{d-1} y + a_{d-2} x^{d-1} y^2 + \cdots + a_1 x y^{d-1} + a_0 y^d = m$$

with the unknowns x, y, where $a_i, m \in \mathbb{Z}$, $d \geq 3$ and the homogeneous polynomial $P(x, y)$ is nonzero and irreducible in $\mathbb{Z}[x, y]$, has only finitely many integral solutions $x, y \in \mathbb{Z}$.

Thue reproved it in the next year in his famous article [41] where he derived it from a theorem on approximation of algebraic numbers by fractions.

Theorem 1.0.2 (Thue's inequality, 1909) *Let $\alpha \in \mathbb{C}$ be an algebraic number with degree $d \in \mathbb{N}$ and let $\varepsilon \in (0, \frac{1}{2})$. Then only finitely many fractions $\frac{p}{q} \in \mathbb{Q}$ satisfy inequality*

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

In other words, there is a constant $c = c(\alpha, \varepsilon) > 0$ such that every fraction $\frac{p}{q} \neq \alpha$ satisfies inequality

$$\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^{1+\varepsilon+d/2}}.$$

We shall prove Theorem 1.0.2 and thus Theorem 1.0.1 in Section 1.2. In Section 1.1 we collect auxiliary results.

Theorem 1.0.2 is deep and difficult when $\alpha \in \mathbb{R}$ and $d \geq 3$, else one easily proves even stronger bounds. If $\alpha \in \mathbb{C}\backslash\mathbb{R}$ then $|\alpha - \frac{p}{q}| \geq \mathrm{Im}(\alpha) > 0$. If $\alpha \in \mathbb{R}$ has degree $d = 1$ then $\alpha = \frac{a}{b} \in \mathbb{Q}$ and for $\frac{p}{q} \neq \alpha$ we have $|\alpha - \frac{p}{q}| \geq \frac{1}{qb}$. If $\alpha \in \mathbb{R}$ has degree $d = 2$ then there is an $\alpha' \in \mathbb{R}$, $\alpha' \neq \alpha$, such that $P(x) = x^2 + ax + b = (x - \alpha)(x - \alpha')$ has rational coefficients a, b. For every $\frac{p}{q} \in \mathbb{Q}$, as α and α' are irrational, $P(\frac{p}{q}) \neq 0$ and thus $|P(\frac{p}{q})| \geq \frac{1}{kq^2}$ where $k \in \mathbb{N}$ is a common denominator of a and b. Hence, denoting $\delta = |\alpha - \alpha'| > 0$, for every fraction $\frac{p}{q}$ we have

$$\left|\alpha - \frac{p}{q}\right| \begin{cases} \geq 1 & \text{if } |\alpha - p/q| \geq 1 \\ = \frac{|P(p/q)|}{|\alpha' - p/q|} > \frac{1}{(1+\delta)kq^2} & \text{if } |\alpha - p/q| < 1, \end{cases}$$

which is stronger than the bound in Theorem 1.0.2 for $d = 2$. This argument easily generalizes and gives *Liouville's inequality* (Liouville, 1844, [24]): if $\alpha \in \mathbb{R}$ is algebraic with degree d then for any fraction $\frac{p}{q} \neq \alpha$,

$$\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^d}$$

where $c > 0$ depends only on α. For $d \geq 3$ this is weaker than *Thue's inequality* in Theorem 1.0.2 and indeed too weak to yield Theorem 1.0.1. As we will see in 2 of Proposition 1.1.5, any strengthening of Liouville's inequality replacing c with a function going to infinity with q gives Theorem 1.0.1. Thue's inequality is easiest such strengthening known.

Theorem 1.0.1 is trivial for $m = 0$, with no integral solution, since by the irreducibility $P(z,1) = \sum_{i=0}^{d} a_i z^i$, $d \geq 2$, does not have rational roots. Equation with general homogeneous polynomial $P(x,y)$ reduces to the case when it is irreducible because if $P = P_1 P_2$ in $\mathbb{Z}[x,y]$ then P_i are homogeneous and $P(p,q) = m$ with $p,q \in \mathbb{Z}$ implies that $P_i(p,q) = m_i$ where m_i is a divisor of m if $m \neq 0$ or $m_i = 0$. Thus $P(x,y) = m$ reduces to finitely many equations $P_1(x,y) = m_1$ with P_1 an irreducible divisor of P in $\mathbb{Z}[x,y]$ and m_1 a divisor of m or $m_1 = 0$. (We recall in part 2 of Proposition 1.1.4 that P is irreducible in $\mathbb{Z}[x,y]$ iff it is irreducible in $\mathbb{Q}[x,y]$.) If P is irreducible with degree $d \leq 2$, the equation $P(x,y) = m$ may have infinitely many integral solutions, for example if $P(x,y) = ax+by$ and (a,b) divides m or if it is the Pell equation $x^2 - ay^2 = 1$ with non-square $a \in \mathbb{N}$. Finally, we remark that the reduction of Theorem 1.0.1 to Theorem 1.0.2 proves more generally finiteness of solutions of equations $P(x,y) = Q(x,y)$ where $Q \in \mathbb{Z}[x,y]$ is any polynomial with degree smaller than $d/2 - 1$, indeed any integral function satisfying a growth condition (part 3 of Proposition 1.1.5).

1.1 Polynomials, algebraic numbers, lemmas of Siegel and Gauss

For $j \in \mathbb{N}_0$ and a polynomial $a = \sum_{i=0}^{n} a_i x^i$ with $a_i \in \mathbb{C}$ we set

$$\|a\| = \max_{0 \leq i \leq n} |a_i| \quad \text{and} \quad D_j a = \sum_{i=0}^{n} \binom{i}{j} a_i x^{i-j} = \frac{a(x)^{(j)}}{j!}.$$

So $\|a\|$ is the largest modulus of a coefficient in a and D_j is the operator of normalized j-th derivative by x. Clearly, $D_j(\alpha a + \beta b) = \alpha D_j a + \beta D_j b$ for every $\alpha, \beta \in \mathbb{C}$ and $a, b \in \mathbb{C}[x]$. D_j increases the norm $\|\cdot\|$ only by an exponential factor (claim 6 below), the factor $\deg(a)!$ obtained without division by $j!$ would be too big in applications.

Proposition 1.1.1 *Let $a = \sum_{i=0}^{m} a_i x^i$ and $b = \sum_{i=0}^{n} b_i x^i$, $a_m b_n \neq 0$, be polynomials with complex coefficients.*

1. *For every $\alpha \in \mathbb{C}$, $|a(\alpha)| \leq (\deg(a)+1)\|a\| \max(1,|\alpha|)^{\deg(a)}$.*
2. *For every $\alpha, \beta \in \mathbb{C}$, $\|\alpha a + \beta b\| \leq |\alpha| \cdot \|a\| + |\beta| \cdot \|b\|$.*

3. *One has* $\|ab\| \le (\deg(a)+1)\cdot\|a\|\cdot\|b\|$.

4. *If* $a = (x-\alpha)^r b$, *where* $\alpha \in \mathbb{C}$ *is nonzero and* $r \in \mathbb{N}$, *then*
$$\|b\| < (\deg(b)+1)(2\max(1,|\alpha|^{-1}))^{\deg(a)}\|a\|.$$

5. *If* $a \in \mathbb{Z}[x]$ *then* $D_j a \in \mathbb{Z}[x]$ *for every* $j = 0, 1, \ldots$.

6. *For every* $j = 0, 1, \ldots$, $\|D_j a\| \le 2^{\deg(a)}\|a\|$.

Proof. 1. We have $|a(\alpha)| \le \sum_{i=0}^m |a_i|\cdot|\alpha|^i \le \|a\| \sum_{i=0}^m |\alpha|^i$ and the estimate follows upon comparing 1 and $|\alpha|$.

2. Now $\|\alpha a + \beta b\| = \max_{0\le i\le m+n} |\alpha a_i + \beta b_i|$ (where $a_i = 0$ for $i > m$ and similarly for b_i). Since $|\alpha a_i + \beta b_i| \le |\alpha|\cdot|a_i| + |\beta|\cdot|b_i| \le |\alpha|\cdot\|a\| + |\beta|\cdot\|b\|$, we get the stated estimate.

3. The coefficient of x^k in ab equals $\sum_{i+j=k} a_i b_j$, a sum with at most $m+1$ nonzero summands. Its modulus is therefore at most $(m+1)\|a\|\cdot\|b\|$, which gives the stated estimate.

4. In the ring of power series $\mathbb{C}[[x]]$ we have equality $b = (x-\alpha)^{-r}a$, where $(x-\alpha)^{-r} = (-\alpha)^{-r}(1-x/\alpha)^{-r} = (-\alpha)^{-r}\sum_{k\ge 0}\binom{k+r-1}{k}(x/\alpha)^k$. Thus $b = (-\alpha)^{-r}\sum_{k=0}^n \binom{k+r-1}{k}(x/\alpha)^k \cdot a$ and the bound follows by part 3, using that $\binom{k+r-1}{k} < 2^{n+r}$ and $n+r=m$.

5 and 6. They follow from the fact that $\binom{i}{j} \in \mathbb{N}$ and $\binom{i}{j} \le (1+1)^i = 2^i$ for $0 \le j \le i$ and $\binom{i}{j} = 0$ for $j > i$. □

A number $\alpha \in \mathbb{C}$ is *algebraic* if $p(\alpha) = 0$ for a nonzero polynomial $p \in \mathbb{Q}[x]$; if no such p exists we say that α is *transcendental*. We may assume that p is monic (has leading coefficient 1) or that $p \in \mathbb{Z}[x]$ but, in general, not both. If p has both properties, is a monic integral polynomial and $p(\alpha) = 0$, we say that α is an *algebraic integer*. The *degree* of an algebraic number α is the minimum degree of a nonzero $p \in \mathbb{Q}[x]$ with $p(\alpha) = 0$. Clearly, α has degree 1 iff $\alpha \in \mathbb{Q}$. The *minimum polynomial* of an algebraic number α with degree d is the unique monic polynomial $p \in \mathbb{Q}[x]$ such that $p(\alpha) = 0$ and $\deg(p) = d$. If α is an algebraic integer then its minimum polynomial has in fact integral coefficients (3 of Proposition 1.1.4).

Proposition 1.1.2 *Let* $\alpha \in \mathbb{C}$ *be an algebraic number.*

1. *For every* $k, l \in \mathbb{Q}$, $k \ne 0$, *the number* $\beta = k\alpha + l$ *is algebraic and has the same degree as* α.

2. *There exists a* $k \in \mathbb{N}$*, a* denominator of α*, such that* $k\alpha$ *is an algebraic integer.*

3. *If* α *is an algebraic integer then so is* $\alpha + l$ *for any* $l \in \mathbb{Z}$*.*

4. *If* α *is an algebraic integer and* $p(\alpha) = 0$ *for a monic polynomial* $p \in \mathbb{Z}[x]$ *with degree* d *then for* $r = 0, 1, \ldots$ *we have expressions*

$$\alpha^r = \sum_{i=0}^{d-1} c_{r,i}\alpha^i \quad \textit{where} \quad c_{r,i} \in \mathbb{Z} \quad \textit{with} \quad |c_{r,i}| \leq (1 + \|p\|)^r.$$

5. *The minimum polynomial of* α *has only simple roots, is irreducible in* $\mathbb{Q}[x]$ *and divides every* $q \in \mathbb{Q}[x]$ *with root* α*.*

6. *Every nonzero and irreducible polynomial* $p \in \mathbb{Q}[x]$ *with degree* d *has* d *distinct roots which are all algebraic numbers with degree* d*.*

7. *If* α *has degree* d *and is an* m*-tuple root,* $m \geq 1$*, of a nonzero polynomial* $q \in \mathbb{Q}[x]$ *then* $\deg(q) \geq md$*.*

Proof. 1. Since $p(\alpha) = 0$ for a nonzero $p \in \mathbb{Q}[x]$, also $q(\beta) = 0$ where $q(x) = p((x - l)/k) \in \mathbb{Q}[x]$ is nonzero and with $\deg(q) = \deg(p)$. The inverse relation $\alpha = k^{-1}\beta - k^{-1}l$ implies equality of degrees of α and β.

2. Let $\alpha^n + \sum_{i=0}^{n-1} a_i\alpha^i = 0$ where $a_i \in \mathbb{Q}$. Any common denominator $k \in \mathbb{N}$ of the a_is is a denominator of α because multiplication by k^n gives $(k\alpha)^n + \sum_{i=0}^{n-1} a_i k^{n-i}(k\alpha)^i = 0$ and $a_i k^{n-i} \in \mathbb{Z}$.

3. Immediate from the fact that if $p \in \mathbb{Z}[x]$ is monic and $l \in \mathbb{Z}$, then (due to the binomial theorem) also $q(x) = p(x - l) \in \mathbb{Z}[x]$ and is monic.

4. Let $p = \sum_{i=0}^{d} a_i x^i$ with $a_i \in \mathbb{Z}$ and $a_d = 1$. For $r = 0$ we set $c_{0,0} = 1$ and $c_{0,i} = 0$ for $0 < i < d$. Replacing α^d with $-\sum_{i=0}^{d-1} a_i\alpha^i$, for $r > 0$ we get by induction

$$\alpha^r = \alpha \cdot \alpha^{r-1} = \sum_{i=0}^{d-1}(c_{r-1,i-1} - c_{r-1,d-1}a_i)\alpha^i, \; c_{r-1,-1} = 0.$$

Thus $c_{r,i} = c_{r-1,i-1} - c_{r-1,d-1}a_i$ and by induction $|c_{r,i}| \leq |c_{r-1,i-1}| + |c_{r-1,d-1}| \cdot |a_i| \leq (1 + \max_{0 \leq j < d} |a_j|)^r$ for every $r = 0, 1, \ldots$ and $0 \leq i < d$.

5. Let p be the minimum polynomial of α. The minimality of its degree with respect to $p(\alpha) = 0$ implies its irreducibility. Division gives $q = ap + b$ with $a, b \in \mathbb{Q}[x]$ and $\deg(b) < \deg(p)$ or b identically zero. Thus $b(\alpha) = 0$ and the latter must occur, p divides q. This division property shows that p is the minimum polynomial of its each root, not just of α. Hence each root of p is simple, cannot be a root of p'.

6 and 7. Follow from the division property in 5. □

Proposition 1.1.3 (Siegel's lemma) *Any system of m homogeneous linear equations with n unknowns*

$$\sum_{j=1}^{n} a_{i,j}x_j = 0,\ 1 \le i \le m,$$

in which $n > m$ and $a_{i,j} \in \mathbb{Z}$ with $|a_{i,j}| \le A$ for every i, j, has an integral solution $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n$ such that not all α_j are zero and $|\alpha_j| \le \lfloor (nA)^{m/(n-m)} \rfloor$ for every j.

Proof. Let $f_i = \sum_{j=1}^{n} a_{i,j}x_j$ be the form in the i-th equation and let $a_i = \sum_{j=1}^{n} \max(0, a_{i,j})$, $b_i = \sum_{j=1}^{n} \min(0, a_{i,j})$. For $r \in \mathbb{N}_0$ there are $(r+1)^n$ tuples $(x_1, \ldots, x_n)$ in the box $\{0, 1, \ldots, r\}^n$ of arguments, and the values $(f_1, \ldots, f_m)$ of the m forms on them fall in the box $\prod_{i=1}^{m}\{b_i r, b_i r+1, \ldots, a_i r\}$ that contains $\prod_{i=1}^{m}(ra_i - rb_i + 1) \le (rnA + 1)^m$ tuples. If $(rnA + 1)^m < (r+1)^n$, by the pigeon-hole principle two distinct n-tuples are mapped by the forms to the same m-tuple. Their difference, which we denote $(\alpha_1, \ldots, \alpha_n)$, is mapped by the forms to the m-tuple of zeros, has $|\alpha_i| \le r$ and not all α_i are zero. We check that $r = \lfloor (nA)^{m/(n-m)} \rfloor$ satisfies the required inequality and are done: since $r+1 > (nA)^{m/(n-m)}$, indeed $(r+1)^n > ((r+1)nA)^m > (rnA + 1)^m$. $\square$

An integral polynomial $\sum_{i=0}^{n} a_i x^i \in \mathbb{Z}[x]$ is *primitive* if its coefficients are together coprime, no integer greater than 1 divides every a_i. Claims 1 and 2 below are known as *Gauss lemma*.

Proposition 1.1.4 *Integral and rational polynomials have the following properties.*

1. *If $a, b \in \mathbb{Z}[x]$ are primitive then so is their product ab.*
2. *If $a \in \mathbb{Z}[x]$ and is primitive, $b \in \mathbb{Q}[x]$ and $ab \in \mathbb{Z}[x]$ then $b \in \mathbb{Z}[x]$ too. Consequently, if $c \in \mathbb{Z}[x]$ is irreducible in $\mathbb{Z}[x]$ then it is irreducible in $\mathbb{Q}[x]$.*
3. *The minimum polynomial of an algebraic integer has integral coefficients.*
4. *If $a \in \mathbb{Z}[x]$ has leading coefficient $l \in \mathbb{N}$ and a root $\frac{p}{q} \in \mathbb{Q}$, $(p, q) = 1$, with multiplicity $m \in \mathbb{N}$ then $q^m \le l$.*

Proof. 1. Let $a = \sum_{i=0}^m a_i x^i$, $b = \sum_{i=0}^n b_i x^i$, p be a prime number and a_k, b_l be the coefficients with least indices that are not divisible by p. The coefficient of x^{k+l} in ab equals $a_k b_l + \sum_{i=0}^{k-1} a_i b_{k+l-i} + \sum_{i=0}^{l-1} a_{k+l-i} b_i$. This is not divisible by p either because $a_k b_l$ is not divisible by p but each summand in the two sums is divisible by p. Thus ab is primitive as no prime divides all its coefficients.

2. We put the coefficients in b in the lowest terms a denote by $k \in \mathbb{N}$ their least common denominator. It follows that $kb \in \mathbb{Z}[x]$ and is primitive. The equality $a \cdot kb = kab$ implies by the previous result that $k = 1$ and so $b \in \mathbb{Z}[x]$. Let $c \in \mathbb{Z}[x]$ and $c = ab$ with $a, b \in \mathbb{Q}[x]$. For appropriate $k \in \mathbb{N}$ is $ka \in \mathbb{Z}[x]$ and is primitive. Then $c = (ka)(k^{-1}b)$ and, as we know, $k^{-1}b \in \mathbb{Z}[x]$ as well. Thus reducibility of c in $\mathbb{Q}[x]$ implies its reducibility in $\mathbb{Z}[x]$.

3. Let $\alpha \in \mathbb{C}$ be an algebraic integer, $q \in \mathbb{Z}[x]$ be monic with $q(\alpha) = 0$ and $p \in \mathbb{Q}[x]$ be the minimum polynomial. Then $q = pa$ for some $a \in \mathbb{Q}[x]$ (by 5 of Proposition 1.1.2). We take $k \in \mathbb{N}$ so that $kp \in \mathbb{Z}[x]$ and is primitive. Equality $q = (kp)(k^{-1}a)$ implies, by 2, that $k^{-1}a \in \mathbb{Z}[x]$. As q is monic, the leading coefficients of kp and $k^{-1}a$ are ± 1, and we see that $k = 1$ and $p \in \mathbb{Z}[x]$.

4. The polynomial $(qx-p)^m = q^m x^m + \cdots + (-p)^m$ is primitive and, as $(qx-p)^m = q^m(x-p/q)^m$, divides $a = lx^n + \cdots + a_0$ in $\mathbb{Q}[x]$, $(qx-p)^m b = a$ for some $b \in \mathbb{Q}[x]$. By part 2, $b \in \mathbb{Z}[x]$ and therefore $q^m k = l$ where $k \in \mathbb{N}$ is the leading coefficient in b. In particular, $q^m \leq l$. □

Proposition 1.1.5 *The following reductions hold.*

1. *If Theorem 1.0.2 holds for every algebraic integer $\alpha \in \mathbb{R}$ with degree at least 3 and $|\alpha| \leq \frac{1}{2}$ then it holds for every algebraic number $\alpha \in \mathbb{C}$.*

2. *Suppose that each algebraic number $\alpha \in \mathbb{C}$ with degree $d \geq 3$ satisfies a strengthened Liouville's inequality, which means that there exist functions $c_\alpha : \mathbb{N}_0 \to \mathbb{R}$ such that $c_\alpha(q) \to +\infty$ as $q \to \infty$ and every fraction $\frac{p}{q}$, $(p,q) = 1$, satisfies $|\alpha - p/q| > c_\alpha(q)/q^d$. Then for every homogeneous, nonzero and irreducible polynomial $P \in \mathbb{Z}[x,y]$ with $\deg(P) \geq 3$ we have $|P(p,q)| \to +\infty$ as $\max(|p|,|q|) \to +\infty$ on $p, q \in \mathbb{Z}$, because, with $m = \max(|p|,|q|)$ for $p, q \in \mathbb{Z}$,*

$$|P(p,q)| \gg_P \min(m^d, \min_{\alpha \in A} c_\alpha(m))$$

where A is the set of roots of the polynomials $P(x,1)$ and $P(1,y)$. In particular, this implies that every Thue equation has only finitely many solutions.

3. *Theorem 1.0.2 implies Theorem 1.0.1 and in fact more strongly that every homogeneous, nonzero and irreducible polynomial $P \in \mathbb{Z}[x,y]$ with degree $d \geq 3$ satisfies for every $p,q \in \mathbb{Z}$ and $\varepsilon > 0$ the inequality*

$$|P(p,q)| \gg_{P,\varepsilon} \max(|p|,|q|)^{d/2-1-\varepsilon}.$$

Proof. 1. As we know, we may restrict to real algebraic numbers with degree ≥ 3. Let $\alpha \in \mathbb{R}$ be algebraic with degree d. Using parts 1, 2 and 3 of Proposition 1.1.2, we select numbers $k \in \mathbb{N}$ and $l \in \mathbb{Z}$ so that $k\alpha + l$ is an algebraic integer with degree d and $|k\alpha + l| \leq \frac{1}{2}$ (so $l = \lfloor k\alpha \rfloor$ or $l = \lceil k\alpha \rceil$). Now if $\frac{p}{q} \neq \alpha$ then $\frac{kp+lq}{q} \neq k\alpha + l$ and $|\alpha - \frac{p}{q}| = \frac{1}{k}|k\alpha + l - \frac{kp+lq}{q}| > (c/k)q^{-1-\varepsilon-d/2}$ for a constant $c > 0$ by the assumption.

2. Let $P(x,y)$ be as stated. We factorize it as

$$P(x,y) = ay^d \textstyle\prod_{i=1}^d (x/y - \alpha_i)$$

where $a \in \mathbb{Z}$ is nonzero and the α_i are roots of $P(x,1)$. Since $P(x,1)$ is irreducible in $\mathbb{Z}[x]$ and in $\mathbb{Q}[x]$ (part 2 of Proposition 1.1.4), the α_i are distinct algebraic numbers with degree d (part 6 of Proposition 1.1.2). Let

$$\delta = \tfrac{1}{2} \min_{i \neq i'} |\alpha_i - \alpha_{i'}| > 0$$

and $p,q \in \mathbb{Z}$ be arbitrary with $m = \max(|p|,|q|)$. If $q = 0$, then $|P(p,q)| = |P(p,0)| \gg |p|^d = m^d$ because $P(x,0)$ has degree d (by irreducibility). Thus we may assume that $q \neq 0$. We may also assume that $|q| \geq |p|$; in the case $|p| \geq |q|$ we use the symmetric factorization obtained by taking out ax^d and the roots of $P(1,y)$. If $|\alpha_i - p/q| < \delta$ for no i then $|P(p,q)| = |aq^d| \prod_{i=1}^d |\alpha_i - p/q| \geq |aq^d|\delta^d \gg |q|^d = m^d$. If, say, $|\alpha_1 - p/q| < \delta$ then triangle inequality implies that $|\alpha_i - p/q| > \delta$ for $i > 1$ and we get, using the strengthened Liouville's inequality for α_1, that $|P(p,q)| > |aq^d| \cdot |\alpha_1 - p/q|\delta^{d-1} \gg c_{\alpha_1}(|q|) = c_{\alpha_1}(m)$. So we get the lower bound on $|P(p,q)|$.

3. Immediate from part 2 and Theorem 1.0.2. □

1.2 Proof of Thue's theorem

We prove Theorem 1.0.2, which by part 3 of Proposition 1.1.5 proves Theorem 1.0.1 as well. We assume that $\alpha \in \mathbb{R}$ is an algebraic integer with degree $d \geq 3$ and $|\alpha| \leq \frac{1}{2}$; part 1 of Proposition 1.1.5 says that Theorem 1.0.2 reduces to this case. We proceed in four steps. First we construct a nonzero bivariate polynomial $F(x, y) = P(x) + yQ(x)$ with integral and not too large coefficients that vanishes at $(x, y) = (\alpha, \alpha)$ to a high order. We derive from this that for any two fractions u, v approximating α the derivatives $(D_j F)(u, v)$ are close to zero. They are, however, not too close to zero since we show that $(D_h F)(u, v) \neq 0$ for not too large h. Finally, assuming existence of infinitely many very close rational approximations to α, we select appropriately two of them, u and v, so that the upper and lower bounds on $|(D_h F)(u, v)|$ from steps 2 and 3 become contradictory.

Proposition 1.2.1 *Let $d, m, n \in \mathbb{N}$, $d \geq 3$, and $\lambda \in (0, \frac{1}{2})$ be such that $m = (2n + 2)(1 - \lambda)/d$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ be an algebraic integer of degree d. Then there exist nonzero integral polynomials $P, Q \in \mathbb{Z}[x]$ such that*

1. $\deg(P), \deg(Q) \leq n$;
2. $\|P\|, \|Q\| \leq c_1^{n/\lambda}$ *where $c_1 > 1$ depends only on α;*
3. $D_j(P(x) + \alpha Q(x))(\alpha) = 0$ *for every $0 \leq j < m$.*

Moreover, P and Q are not proportional ($P(x)/Q(x)$ is not constant).

Proof. In accordance with condition 1 we write $P(x) = \sum_{i=0}^{n} a_i x^i$, $Q(x) = \sum_{i=0}^{n} b_i x^i$ with a_i, b_i being $2n + 2$ unknown coefficients. The vanishing of derivatives in condition 3 means that $\sum_{i=0}^{n} \binom{i}{j}(a_i \alpha^{i-j} + b_i \alpha^{i-j+1}) = 0$ for $0 \leq j < m$ (recall that $\binom{i}{j} = 0$ for $j > i$). Replacing the powers of α by the expressions of part 4 of Proposition 1.1.2 gives us m equalities

$$\textstyle\sum_{k=0}^{d-1} \alpha^k \sum_{i=j}^{n} \binom{i}{j}(c_{i-j,k} a_i + c_{i-j+1,k} b_i) = 0, \ 0 \leq j < m.$$

Here $c_{r,k} \in \mathbb{Z}$ with $|c_{r,k}| < c_0^r$ and $c_0 > 1$ depends only on α. These equalities are satisfied if (and only if) all dm coefficients of α^k, $0 \leq k < d$, are zero. This gives dm homogeneous linear equations with $2n + 2$ unknowns a_i, b_i and integral coefficients $\binom{i}{j} c_{r,k}$, $j \leq i \leq n$, $0 \leq r \leq n$ and $0 \leq k < d$, whose absolute values are bounded by $A = (2c_0)^n$. Since $2n + 2 > dm$, by Siegel's lemma there exist $a_i, b_i \in \mathbb{Z}$, not all zero, satisfying these equations and bounded by $|a_i|, |b_i| \leq (2n+2)A^{md/(2n+2-md)} < (2n+2)A^{1/\lambda} \leq (8c_0)^{n/\lambda}$, which is the bound required in condition 2, with $c_1 = 8c_0$.

These a_i, b_i give polynomials P, Q that are not both zero and satisfy conditions 1–3. We show that they are in fact both nonzero and even non-proportional. Suppose for contrary that $Q \neq 0$ but $P = cQ$ for a constant c (possibly zero); the case when $P \neq 0$ and $Q = cP$ is similar. By 3, the polynomial $R(x) = (c+\alpha)Q(x)$ has at $x = \alpha$ zero of order at least m. It is a nonzero polynomial because $Q \neq 0$ and $c + \alpha \neq 0$ as $c \in \mathbb{Q}$. Thus $Q(x) = (c+\alpha)^{-1}R(x)$ has at $x = \alpha$ zero of order at least m. By part 7 of Proposition 1.1.2, $\deg(Q) \geq md = (2n+2)(1-\lambda) > n+1$, which contradicts $\deg(Q) \leq n$. $\square$

Proposition 1.2.2 *Let d, m, n, λ, α and the corresponding polynomials P, Q and constant c_1 be as in Proposition 1.2.1 and $u = \frac{p}{q}, v = \frac{r}{s}$ be two fractions satisfying $q, s \geq 2$, $|\alpha - u| < q^{-\mu}$ and $|\alpha - v| < s^{-\mu}$ where $\mu > 1$. Then for every $0 \leq j < m$,*

$$|D_j(P(x) + vQ(x))(u)| \leq c_2^{n/\lambda}(q^{-\mu(m-j)} + s^{-\mu})$$

with $c_2 > 1$ depending only on α.

Proof. Let $F(x,y) = P(x) + yQ(x)$. Since $F(x,\alpha)$ has at $x = \alpha$ zero of order at least m, we have $F(x,y) = F(x,\alpha) + (y-\alpha)Q(x) = (x-\alpha)^m R(x) + (y-\alpha)Q(x)$ where $R \in \mathbb{C}[x]$. From this we get $D_jF(x,y) = (x-\alpha)^{m-j}S(x) + (y-\alpha)D_jQ(x)$ for some $S \in \mathbb{C}[x]$. Using parts 1 and 2 of Proposition 1.1.1 and the fact that $|u|, |v| < 1$ (since $|\alpha| < \frac{1}{2}$), we get

$$\begin{aligned}
|D_j(P(x) + vQ(x))(u)| &= |D_jF(x,y)(u,v)| \\
&= |(u-\alpha)^{m-j}S(u) + (v-\alpha)D_jQ(u)| \\
&\leq q^{-\mu(m-j)}|S(u)| + s^{-\mu}|D_jQ(u)| \\
&\leq q^{-\mu(m-j)}(n+1)\|S\| + s^{-\mu}(n+1)\|D_jQ\|.
\end{aligned}$$

Now $\|D_jQ\| \leq (2c_1)^{n/\lambda}$ by part 6 of Proposition 1.1.1 and 2 of Proposition 1.2.1. Equality $D_jF(x,\alpha) = (x-\alpha)^{m-j}S(x)$ gives, by part 4 of Proposition 1.1.1, that $\|S\| < (\deg(S)+1)(2/|\alpha|)^{n-j}\|D_jF(x,\alpha)\| \leq (4/|\alpha|)^n(4c_1)^{n/\lambda}$ because $\deg(S) \leq n < 2^n$, $|\alpha| < 1$ and both $\|D_jP\|, \|D_jQ\|$ are bounded by $(2c_1)^{n/\lambda}$. Hence $\|S\| \leq (16c_1/|\alpha|)^{n/\lambda}$. In view of these bounds and $2(n+1) \leq 4^n$ we obtain the stated estimate with $c_2 = 64c_1/|\alpha|$. $\square$

Proposition 1.2.3 *Let $d, m, n, \lambda, \alpha, P, Q, u = \frac{p}{q}$ and $v = \frac{r}{s}$ be as before, in Proposition 1.2.2. Then $D_h(P(x)+vQ(x))(u) \neq 0$ for some $h \in \mathbb{N}$ satisfying $h \leq 1 + (c_3/\lambda)n/\log q$ where $c_3 > 0$ depends only on α.*

Proof. The integral polynomial $W = P'Q - PQ'$ is nonzero because by Proposition 1.2.1 P, Q are non-proportional (so $(P/Q)' = W/Q^2$ is nonzero). Using the binomial formula (for derivatives of a product) we get that $W^{(j)} = \sum_{i=0}^{j} \binom{j}{i}(P^{(i+1)}Q^{(j-i)} - P^{(j-i)}Q^{(i+1)})$ for every $j \in \mathbb{N}_0$. Let $h \in \mathbb{N}_0$ be the minimum number with $(P(x) + vQ(x))^{(h)}(u) \neq 0$, which exists because $P(x) + vQ(x) \neq 0$. So $P^{(j)}(u) + vQ^{(j)}(u) = 0$ for $0 \leq j < h$. Elimination of v gives equalities $P^{(j)}(u)Q^{(i)}(u) - P^{(i)}(u)Q^{(j)}(u) = 0$ for $0 \leq i, j < h$. Hence $W^{(j)}(u) = 0$ for $0 \leq j < h - 1$ and W has at $x = u = \frac{p}{q}$ zero of order at least $h - 1$. By part 4 of Proposition 1.1.4, $q^{h-1} \leq \|W\|$. By the estimates in Proposition 1.1.1 and 2 of Proposition 1.2.1, $\|W\| \leq 2n\|PQ\| \leq 2n(n+1)c_1^{2n/\lambda} \leq (4c_1^2)^{n/\lambda}$. Thus the stated bound on h holds with $c_3 = \log(4c_1^2)$. $\square$

We prove by contradiction Theorem 1.0.2. We are given an algebraic integer $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ of degree $d \geq 3$ and an $\varepsilon \in (0, \frac{1}{2})$, and we assume that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{1+\varepsilon+d/2}}$$

holds for infinitely many $\frac{p}{q} \in \mathbb{Q}$. We derive a contradiction.

We fix an even $t \in \mathbb{N}$ so that $\lambda = 2/(t+2) < \varepsilon/2d$, thus $0 < \lambda < \frac{1}{12}$ and $t \geq 24$, and let n run through the arithmetic progression $n = i(t/2+1)d - 1$, $i = 1, 2, \ldots$, thus $m = (2n+2)(1-\lambda)/d = it$. Let c be the maximum of the constants $c_1^{1/\lambda}, c_2^{1/\lambda}$ and c_3/λ of Propositions 1.2.1–1.2.3; c depends only on α and ε. We set

$$\mu = 1 + \varepsilon + d/2 \quad \text{and} \quad \delta = (1 + 2\varepsilon/d)(1 - \lambda) - 1.$$

Note that $\frac{1}{3} > \delta > \varepsilon/2d > 0$. From the infinitely many close rational approximations to α we select two, $u = \frac{p}{q}$ and $v = \frac{r}{s}$, so that $(p, q) = (r, s) = 1$, $2 \leq q < s$, $|\alpha - u| < q^{-\mu}$, $|\alpha - v| < s^{-\mu}$,

$$\log q > 2cd\mu/\delta \quad \text{and} \quad \log s/\log q > t + 2(\mu + t)/\delta.$$

We show that the bounds of Propositions 1.2.2 and 1.2.3 are for u and v contradictory.

We take the unique $m = it$, $i \in \mathbb{N}$, satisfying

$$\frac{\log s}{\log q} - t \le m < \frac{\log s}{\log q}$$

and the corresponding $n = i(t/2+1)d - 1$. We take the polynomials P, Q ensured by Proposition 1.2.1 that correspond to α, d, m, n, λ and take the minimum $h \in \mathbb{N}_0$ such that

$$w = D_h(P(x) + vQ(x))(u) = (D_hP)(u) + v(D_hQ)(u) \neq 0.$$

By the lower bound on m and $\log s/\log q$ we have $m > 6t > 100$. Since $4n/d \ge 2(n+1)/d > m > 100$, we have that $n > 25d$. By Proposition 1.2.3, $n > 2d$ and the lower bound on $\log q$ we get that $h < m$ because $h \le 1 + cn/\log q < 1 + n/2d < n/d < \frac{11}{6}(n+1)/d < (2n+2)(1-\lambda)/d = m$. We have inequalities

$$(q^{n-h}s)^{-1} \le |w| < c^n(q^{-\mu(m-h)} + s^{-\mu}) \le (2c)^n q^{-\mu(m-h)}.$$

The lower bound follows from the fact that $0 \neq w \in \mathbb{Q}$ and $q^{n-h}sw \in \mathbb{Z}$ because $D_hP, D_hQ \in \mathbb{Z}[x]$ and have degrees at most $n-h$. The upper bound follows from Proposition 1.2.2 and the fact that $s > q^m$ (by the second inequality defining m). Taking logarithms we get

$$\mu m - \mu h + h - n \le \frac{\log s}{\log q} + \frac{n\log(2c)}{\log q} \le m + t + \frac{n\log(2c)}{\log q},$$

by the first inequality defining m. Using the bound on h from Proposition 1.2.3 we get

$$(\mu-1)m - n \le \mu h + t + \frac{n\log(2c)}{\log q} \le \mu(1 + cn/\log q) + t + \frac{n\log(2c)}{\log q}.$$

Now $(\mu-1)m - n > (\varepsilon + d/2)2n(1-\lambda)/d - n = \delta n$. (Here we see that the $d/2$ in $\mu = 1 + \varepsilon + d/2$ is best possible in this argument.) By the lower bound on $\log q$ we have

$$\delta n \le \mu + \delta n/4 + t + \delta n/4 = \mu + t + \delta n/2$$

and

$$n \le 2(\mu + t)/\delta.$$

This is a contradiction because for large $i \in \mathbb{N}$ is $n = i(t/2+1)d - 1$ greater than any bound depending only on α and ε.

1.3 Remarks

The proof of Thue's theorem in Section 1.2 follows the excellent exposition of Zannier [46, chapter 2]. I was fascinated by it ever since I read 20^+ years ago the condensed (and then to me quite enigmatic) section in Sprindžuk [38, chapter 1.2]. We briefly mention three developments connected to Thue's theorem.

Strengthenings of Thue's inequality. In 1921 C. L. Siegel [36] improved the exponent $1 + \varepsilon + d/2$ to $\varepsilon + 2\sqrt{d}$ (more precisely, to $\varepsilon + \min_{s=1,2,\dots} s + d/(s+1)$) and then, independently, F. J. Dyson [10] and A. O. Gel'fond [16], [15] to $\varepsilon + \sqrt{2d}$. In 1955 K. Roth [32] achieved the ultimate improvement to $\varepsilon + 2$. Roth's theorem states:

- *For every algebraic number $\alpha \in \mathbb{C}$ and every $\varepsilon > 0$ only finitely many fractions $\frac{p}{q}$ satisfy the inequality $|\alpha - \frac{p}{q}| < q^{-\varepsilon-2}$.*

In 1958 was Roth awarded for his result Fields medal. The proof of Roth's theorem can be found, for example, in Bombieri and Gubler [7], Schmidt [33] or Steuding [39].

Finiteness of solution sets of binary Diophantine equations. Siegel used his strengthening of Thue's inequality as a tool for his celebrated theorem on integral points on algebraic curves. Siegel's theorem [37] says:

- *Write the equation $F(x, y) = 0$ with nonzero and irreducible polynomial $F \in \mathbb{C}[x, y]$ as $P(x, y) = Q(x, y)$ where P is nonzero homogeneous and $\deg(P) > \deg(Q)$. If the equation has infinitely many integral solutions $x, y \in \mathbb{Z}$ then $P(z, 1)$ has at most two distinct roots and there are nonconstant rational functions a, b in $\mathbb{C}(t)$ such that $F(a(t), b(t)) = 0$ identically.*

An algorithm can be based on Siegel's theorem that for every binary Diophantine equation determines whether it has infinitely many integral solutions ([6], [46]). But it is open if an algorithm exists determining whether the solution set is nonempty. For a general Diophantine equation in two or more unknowns the last problem is undecidable and no such algorithm exists (by the Davis–Putnam–Robinson–Matijasevič theorem of 1970, [8] and [26], solving the tenth problem of Hilbert). For the proof of Siegel's theorem see Bombieri and Gubler [7] or Hindry and Silverman [21].

Effective solution of Thue equation. Both Thue's inequality with its strengthenings and Siegel's theorem are non-effective, their proofs do not provide algorithms for determining solution sets. Could some or all Thue equations be effectively solved? In 1918 Thue himself obtained such results in a pioneering but often misinterpreted article [42], [43, p. 565–571] where he described effective resolution of certain binomial equations $ax^n - by^n = m$. For example, it follows from his general theorem that if $m, x, y \in \mathbb{Z}$ are such that $x^7 - 17y^7 = m$ then $\max(|x|, |y|) \leq 700|m|^4$. Similar result was obtained in 1964 by A. Baker [2] who showed, for example, that if $m, x, y \in \mathbb{Z}$ are such that $x^3 - 2y^3 = m$ then $\max(|x|, |y|) \leq (300000|m|)^{23}$. Shortly later, in 1966–68, in a breakthrough [3] (awarded by Fields medal in 1970) he devised a method giving, among other applications, an effective bound on solutions to any Thue equation. Baker's effective version of Thue's theorem [4, chapter 4] tells us:

- *Let $\alpha_1, \dots, \alpha_n, \mu$ be algebraic integers of a number field K with degree d such that all α_i are distinct, $n \geq 3$, $\mu \neq 0$ and $K = \mathbb{Q}(\alpha_1, \dots, \alpha_n, \mu, \beta)$ for an algebraic number β. Then every solution to*

  $$(x - \alpha_1 y)(x - \alpha_2 y) \dots (x - \alpha_n y) = \mu$$

  in algebraic integers x, y of K satisfies

  $$\max(h(x), h(y)) < \exp((dH)^{(10d)^5})$$

  where $h(\cdot)$ is the size of algebraic numbers and H is the largest absolute value of a coefficient in the defining polynomials of $\alpha_1, \dots, \alpha_n, \mu$ and β.

(Defining polynomial of an algebraic number α with degree d is the unique nonzero polynomial $p \in \mathbb{Z}[x]$ with degree d such that $p(\alpha) = 0$, p has coprime coefficients and positive leading coefficient. For the definition of number fields etc. see Section 3.1.)

Chapter 2

Dirichlet's theorem on primes in arithmetic progression

Nun läßt sich die Methode, durch die ich den Tschebyschefschen Satz, laut dessen es zwischen ξ und 2ξ stets wenigstens eine Primzahl gibt, bewiesen habe, auf den Fall der obenerwähnten arithmetischen Reihen übertragen.

P. Erdős [11][1]

In 1837, Peter L. Dirichlet (1805–1859) [9] founded analytic number theory by proving this fundamental result.

Theorem 2.0.1 (Dirichlet, 1837) *For every two coprime numbers $a, m \in \mathbb{N}$, the arithmetic progression $a, a+m, a+2m, a+3m, \ldots$ contains infinitely many prime numbers.*

[1] "Then, the method, by which I have proved Chebyshev's theorem saying that between ξ and 2ξ there is at least one prime number, allows be carried over to the case of aforementioned arithmetic progressions." Erdős refers to his work P. Erdös, Beweis eines Satzes von Tschebyschef, Acta Litt. ac Scient. Regiae Univ. Hungaricae Francisco Josephinae **5** (1930–1932), 194–198.

In [22] we presented its classical proof, which goes back to Dirichlet and uses Dirichlet's L-functions and the estimate

$$\sum_p \frac{\chi(p)}{p^s} = O(1),\ s \to 1^+,$$

where χ is a non-principal Dirichlet character modulo m. Infinite sums, infinite products and logarithmic function in complex domain are indispensable in it. Here we present two different proofs, one complete and the other partial. The first proof in Section 2.2, due to H. N. Shapiro (1922), uses only finite sums, with the exception of $\sum_{n=1}^{\infty} \frac{\chi(n)}{n}$, no L-functions and no complex function theory. It is based on the estimate (Propositions 2.2.1 and 2.2.2)

$$\sum_{n<x} \frac{\chi(n)\Lambda(n)}{n} = O(1),\ x \to +\infty,$$

where χ is as before and Λ denotes the von Mangoldt function. In Section 2.1 we collect all results needed for the proof.

The second, partial, proof in Section 2.3 is due to P. Erdős (1913–1996) and is even more elementary—no complex numbers now—but works for finitely many moduli m only, those with

$$\sigma(m) = \sum_{p<m,\ p \nmid m} \frac{1}{p} < 1$$

and for every modulus dividing such m. For example, it works for moduli $m = 5$ and 6 as $\sigma(5) = \frac{5}{6} < 1$ and $\sigma(6) = \frac{1}{5} < 1$ and also for $m = 7$ because although $\sigma(7) = \frac{31}{30} > 1$, we fortunately have $\sigma(14) = \frac{1504}{2145} < 1$. It works for every $m = 1, 2, \ldots, 28$ but not for $m = 29$ (because $\sigma(29m) > 1$ for every $m \in \mathbb{N}$). The complete list of $m \in \mathbb{N}$ with $\sigma(m) < 1$ follows:

$1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 26, 28, 30, 36, 40, 42, 48, 50, 54, 60, 66, 70, 72, 78, 84, 90, 96, 102, 108, 114, 120, 126, 132, 138, 150, 156, 168, 180, 210, 240, 270, 300, 330, 390, 420, 630, 840.$

To get an idea of the proof recall Erdős's proof of Chebyshev's lower bound on the prime counting function $\pi(x)$. The number $\binom{2n}{n} = \frac{(n+1)(n+2)\ldots 2n}{1\cdot 2\cdot \ldots \cdot n}$ is $> 2^{2n+o(n)}$ and its prime factorization has only low powers: if p^k divides $\binom{2n}{n}$ then $p^k \le 2n$. Thus $(2n)^{\pi(2n)} > 2^{2n+o(n)}$ and $\pi(2n) > (\log 2 + o(1))2n/\log(2n)$. In the case of Dirichlet's theorem Erdős constructs fractions $Q_n(a,m) > 0$ with similar properties: $Q_n(a,m)$ is exponentially big in n if $\sigma < 1$ and any prime power p^k, $k \in \mathbb{Z}$, in its factorization with $k \geq 0$ is small and moreover most of these powers have $k = 1$ and $p \equiv a$ modulo m. For $\sigma < 1$ and $n \to \infty$, the infinitude of primes $p \equiv a$ modulo m follows. This is an ideal elementary proof of Dirichlet's theorem as it should be, with the blemish that it does not work for almost all m...

2.1 Characters, Abel's lemma and summation, Möbius function and von Mangoldt function

A *character* of a finite Abelian group $G = (G, \cdot)$ is a homomorphism

$$\chi :\ G \to \mathbb{C}^\times$$

to the group $(\mathbb{C}^\times, \cdot)$ of nonzero complex numbers, $\chi(ab) = \chi(a)\chi(b)$ for every pair of elements of G. It follows that $\chi(1_G) = 1$ and $\chi(a)^n = 1$ for every $a \in G$ where $n = |G|$ is the order of G. The values of χ belong to the n-th roots of unity and $|\chi(a)| = 1$. The *principal character*, denoted χ_0, has all values equal to 1. The set of all characters of G, denoted G^*, is a group with respect to pointwise multiplication: the product of $\psi, \chi \in G^*$ is the character $\psi\chi$ with values

$$(\psi\chi)(a) = \psi(a)\chi(a).$$

In this group $1_{G^*} = \chi_0$ and inverses are obtained by complex conjugation, $\chi^{-1}(a) = \chi(a)^{-1} = \overline{\chi(a)}$, since $|\chi(a)| = 1$. Also, $\chi^{-1}(a) = \chi(a^{-1})$. Instead of χ^{-1} we write $\overline{\chi}$. We will need only characters of the group

$$G(m) = ((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$$

of residues modulo $m \in \mathbb{N}$ coprime to m but it is convenient to derive their properties in general setting. Recall that $G(m)$ has order

$$\varphi(m) = m(1 - p_1^{-1})(1 - p_2^{-1}) \ldots (1 - p_r^{-1})$$

where the p_i are prime divisors of m. Euler's function $\varphi(m)$ counts the numbers among $1, 2, \ldots, m$ that are coprime to m.

Proposition 2.1.1 *Let G, H be finite Abelian groups and $a, b \in G$.*

1. *If $G \subset H$ and H/G is a cyclic group of order n then every character of G has exactly n extensions to a character of H. Consequently, $|H^*| = n|G^*|$.*

2. *For every G, $|G^*| = |G|$. For every $a \in G$, $a \neq 1_G$, there is a character χ of G such that $\chi(a) \neq 1$.*

3. *The sum $\sum_{a \in G} \chi(a)$ equals $|G|$ if $\chi = \chi_0$ and is 0 if χ is non-principal.*

4. *The sum $\sum_{\chi \in G^*} \chi(a)$ equals $|G^*| = |G|$ if $a = 1_G$ and is 0 if $a \neq 1_G$.*

5. *The sum $\sum_{\chi \in G^*} \overline{\chi}(a)\chi(b)$ equals $|G^*| = |G|$ if $a = b$ and is 0 if $a \neq b$.*

Proof. 1. Let aG be a generator of H/G. This means that $a^n = b \in G$ and every $c \in H$ has a unique representation $c = a^i g$ with $0 \leq i < n$ and $g \in G$. Let $\chi \in G^*$. We show that the extensions of χ to characters of H 1-1 correspond to the n-th roots α of the number $\chi(b)$. If $\psi \in H^*$ extends χ then $\psi(c) = \psi(a^i g) = \psi(a)^i \chi(g)$ and, as $\psi(a)^n = \psi(a^n) = \chi(b)$, $\psi(a) = \alpha$ for some α. Thus we define n distinct mappings $\psi = \psi_\alpha : H \to \mathbb{C}^\times$ by $\psi(c) = \alpha^i \chi(g)$ where $c = a^i g$. These are extensions of χ and we only have to check that they are characters of H. If $c_1 = a^{i_1} g_1$ and $c_2 = a^{i_2} g_2$ are elements of H then $\psi(c_1 c_2) = \psi(a^{i_1+i_2} g_1 g_2)$. If $i_1 + i_2 < n$ then by the definition of ψ the last value equals $\alpha^{i_1+i_2}\chi(g_1 g_2) = \alpha^{i_1}\alpha^{i_2}\chi(g_1)\chi(g_2) = \psi(c_1)\psi(c_2)$. Else $n \leq i_1 + i_2 < 2n$ and $a^{i_1+i_2} = a^i b$ where $i = i_1 + i_2 - n$. Then again $\psi(c_1 c_2) = \alpha^i \chi(b g_1 g_2) = \alpha^i \chi(b)\chi(g_1)\chi(g_2) = \alpha^{i+n}\chi(g_1)\chi(g_2) = \alpha^{i_1}\alpha^{i_2}\chi(g_1)\chi(g_2) = \psi(c_1)\psi(c_2)$. To see that $|H^*| = n|G^*|$, note that every $\psi \in H^*$ is an extension of its restriction $\psi \,|\, G \in G^*$.

2. It is simple to obtain a chain of subgroups $\{1_G\} = G_0 \subset G_1 \subset \ldots \subset G_k = G$ such that every factor G_{i+1}/G_i is a cyclic group with order n_{i+1}; for every proper subgroup G_i and every $g \in G \backslash G_i$, $G_{i+1} = \langle G_i \cup \{g\} \rangle$ is a proper cyclic extension of G_i. By part 1, $|G^*| = n_k |G^*_{k-1}| = \cdots = n_k n_{k-1} \ldots n_1$ as $|G^*_0| = 1$. But also $|G| = \prod_{i=0}^{k-1} |G_{i+1}/G_i| = n_k n_{k-1} \ldots n_1$ and $|G^*| = |G|$.

For a given $a \in G$ distinct from 1_G we start the chain of cyclic extensions with $G_1 = \langle a \rangle$. By part 1, $|G^*_1| = |G_1| = n_1 \geq 2$ and every $\psi \in G^*_1$ extends via the chain to a $\chi \in G^*$. Since G_1 is a cyclic group generated by a, its characters are determined by their values on a and there is a $\psi \in G^*_1$ with $\psi(a) \neq 1$. Its extension to a character of G gives the required χ.

3. The first claim is clear as $\chi_0(a) = 1$ for every a. If $\chi \neq \chi_0$, we take $b \in G$ with $\chi(b) \neq 1$. Since $\sum_{a \in G} \chi(a) = \sum_{a \in G} \chi(ab) = \chi(b) \sum_{a \in G} \chi(a)$, the sum equals 0.

4. The first claim is again clear as $\chi(1_G) = 1$ for every χ. If $a \neq 1_G$, we know by part 2 that $\psi(a) \neq 1$ for some $\psi \in G^*$. As in part 3, $\sum_{\chi \in G^*} \chi(a) = \sum_{\chi \in G^*} (\psi\chi)(a) = \psi(a) \sum_{\chi \in G^*} \chi(a)$ implies that the sum equals 0.

5. This follows from part 4 because $\overline{\chi}(a)\chi(b) = \chi(a^{-1}b)$. $\square$

By a *Dirichlet character modulo* $m \in \mathbb{N}$ we mean a mapping

$$\chi : \mathbb{Z} \to \mathbb{C}$$

such that (i) $\chi(a) \neq 0$ iff a is coprime to m, (ii) $\chi(a+m) = \chi(a)$ (periodicity modulo m) and (iii) $\chi(ab) = \chi(a)\chi(b)$ (complete multiplicativity). *Principal Dirichlet character* χ_0 *modulo* m is the characteristic function of numbers coprime to m.

Proposition 2.1.2 *Let χ be a Dirichlet character modulo m and $a, b \in \mathbb{Z}$.*

1. *If $(a, m) > 1$ then $\chi(a) = 0$ else $|\chi(a)| = 1$.*

2. *If $\chi \neq \chi_0$ then $\chi(a) + \chi(a+1) + \cdots + \chi(a+m-1) = 0$ and*

$$|\chi(a) + \chi(a+1) + \cdots + \chi(a+b)| \leq \varphi(m) - 1, \ b \geq 0.$$

3. *If $a \in \mathbb{N}$ and χ has only real values then $\sum_{d \mid a} \chi(d) \geq 0$ and is at least 1 if a is a square.*

4. *There are exactly $\varphi(m)$ Dirichlet characters modulo m.*

5. *If $(a, m) = 1$ then the sum $\sum_{\psi} \overline{\psi}(a)\psi(b)$ over all Dirichlet characters modulo m equals $\varphi(m)$ if $a \equiv b \bmod m$ and is 0 if $a \not\equiv b \bmod m$.*

Proof. There is an obvious 1-1 correspondence between $G(m)^*$ and Dirichlet characters modulo m that preserves principality. If $\chi \in G(m)^*$, we define $\chi' : \mathbb{Z} \to \mathbb{C}$ by $\chi'(a) = 0$ if $(a, m) > 1$ and $\chi'(a) = \chi(a \bmod m)$ if $(a, m) = 1$. It is straightforward to check that χ' has properties (i)–(iii). Conversely, for a Dirichlet character χ' modulo m we define $\chi : G(m) \to \mathbb{C}^\times$ by $\chi(g) = \chi'(a)$ where $a \in \mathbb{Z}$ is an arbitrary number of the congruence class g modulo m. By (iii), $\chi \in G(m)^*$.

1. Clear since $\chi(a)$ is for $(a, m) = 1$ the value of some character of $G(m)$.

2. We have $\sum_{i=a}^{a+m-1} \chi(i) = \sum_{i \in Z} \chi(i) + \sum_{i \in R} \chi(i)$ where Z is a complete system of $\varphi(m)$ residues modulo m coprime to m and R is the rest of the interval. The first sum is 0 by part 3 of Proposition 2.1.1 and in the

second each summand is 0. To get a bound for arbitrary interval, we split it into disjoint intervals I_j with length m and a residual interval R with length $0 \le |R| < m$. We have just shown that $\sum_{i \in I_j} \chi(i) = 0$ and therefore $\sum_{i=a}^{a+b} \chi(i) = \sum_{i \in R} \chi(i)$. If R contains $\varphi(m)$ numbers coprime to m, the last sum is 0 as well. Else it consists besides zeros of at most $\varphi(m) - 1$ summands with modulus 1, which gives the stated bound.

3. If $a = p_1^{e_1} p_2^{e_2} \dots p_r^{e_r}$, complete multiplicativity of χ shows that the sum equals $\prod_{i=1}^{r}(1 + \chi(p_i) + \chi(p_i)^2 + \cdots + \chi(p_i)^{e_i})$. Since $\chi(p_i)$ is $1, -1$ or 0, the factors are correspondingly $e_i + 1$, $\frac{1}{2}(1 + (-1)^{e_i})$ or 1 and are all nonnegative. If each e_i is even, each factor is ≥ 1 and so is their product.

4. Immediate by the described correspondence and part 2 of Proposition 2.1.1.

5. For $(b, m) = 1$ it is a translation of part 5 of Proposition 2.1.1 and for $(b, m) > 1$ it holds trivially. $\square$

We will bound moduli of finite or infinite sums $\sum_i a_i b_i$ for $a_i = \chi(i)$ with non-principal character and $b_i = f(i)$ with a nonnegative and nonincreasing function by means of the next *Abel's lemma.*

Lemma 2.1.3 *Let $a_1, a_2, \dots, a_n$ be complex numbers and $b_1 \ge b_2 \ge \cdots \ge b_n \ge 0$ be real numbers. Then*

$$|a_1 b_1 + a_2 b_2 + \cdots + a_n b_n| \le b_1 \max_{i=1,\dots,n} |a_1 + a_2 + \cdots + a_i|.$$

Proof. We set $A_i = a_1 + a_2 + \cdots + a_i$, $A_0 = b_{n+1} = 0$ and transform the sum as $\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n}(A_i - A_{i-1}) b_i = \sum_{i=1}^{n} A_i (b_i - b_{i+1})$. Thus $|\sum_{i=1}^{n} a_i b_i| \le \sum_{i=1}^{n} |A_i|(b_i - b_{i+1}) \le \max_i |A_i| \cdot \sum_{i=1}^{n}(b_i - b_{i+1}) = b_1 \max_i |A_i|$. $\square$

More precise asymptotics required for the proof of $L(1, \chi) \neq 0$ will be obtained by *Abel's summation.*

Lemma 2.1.4 *Let $a_1, a_2, \dots$ be complex numbers, $A(t) = \sum_{i \le t} a_i$ for $t > 0$ and $f : [1, x] \to \mathbb{R}$, $x > 1$, be a real function that has on the interval first derivative. Then*

$$\sum_{i \le x} a_i f(i) = A(x) f(x) - \int_1^x A(t) f'(t)\, dt.$$

Proof. Let $n = \lfloor x \rfloor$. We transform the sum to $\sum_{i=1}^{n} A(i)(f(i) - f(i+1))$ where f was extended to have derivative on $[1, n+1]$ and value $f(n+1) =$

0. The last sum equals $-\sum_{i=1}^n A(i)\int_i^{i+1} f' = -\sum_{i=1}^n \int_i^{i+1} A(t)f'(t)\ dt = -\int_x^{n+1} A(t)f'(t)\ dt - \int_1^x A(t)f'(t)\ dt = A(x)f(x) - \int_1^x A(t)f'(t)\ dt.$ □

In the proof of $L(1,\chi) \neq 0$ we will also use the following exchange of sum and integral.

Lemma 2.1.5 *Let $f:\ \mathbb{N}^2 \to \mathbb{C}$ and $g:\ [1,x] \to \mathbb{R}$, $x > 1$, g Riemann-integrable on $[1,x]$, be two functions. Then*

$$\int_1^x \Bigg(\sum_{i \le t,\ j \le x/t} f(i,j)\Bigg) g(t)\ dt = \sum_{ij \le x} f(i,j) \int_i^{x/j} g(t)\ dt.$$

Proof. For $t \in [1,x]$ let $R(t) = \{(i,j) \in \mathbb{N}^2 \mid i \le t,\ j \le x/t\}$. We have a unique splitting $[1,x] = I_1 \cup I_2 \cup \cdots \cup I_r$ into closed intervals with disjoint interiors and such that $R(t) = R_k$ is constant for $t \in I_k$ but $R_k \neq R_{k+1}$. The integral on the left then equals $\sum_{k=1}^r \sum_{(i,j)\in R_k} f(i,j) \int_{I_k} g(t)\ dt$. Exchanging summations we get $\sum_{ij\le x} f(i,j) \sum_{k,\ (i,j)\in R_k} \int_{I_k} g(t)\ dt$. The inner sum equals to the inner integral in the displayed formula because the union of the intervals I_k with $(i,j) \in R_k$ is exactly the set $\{t \in [1,x] \mid (i,j) \in R(t)\}$ which equals to the interval $[i, x/j]$. □

Recall that the *Möbius function*

$$\mu:\ \mathbb{N} \to \{-1, 0, 1\}$$

is defined by $\mu(1) = 1$, $\mu(n) = (-1)^r$ if $n = p_1 p_2 \dots p_r$ is a square-free number and $\mu(n) = 0$ else. Part 2 of the next proposition is called *Möbius inversion.*

Proposition 2.1.6 *The Möbius function μ has the following properties.*

1. *For $n \in \mathbb{N}$ the sum $\sum_{d\,|\,n} \mu(d)$ equals 1 if $n = 1$ and is 0 otherwise.*
2. *If $f, g:\ \mathbb{N} \to \mathbb{C}$ are two functions such that $f(n) = \sum_{d\,|\,n} g(d)$ for every n then $g(n) = \sum_{d\,|\,n} \mu(d) f(n/d)$ for every n.*

Proof. 1. The case $n = 1$ is trivial and we assume that $n > 1$. The prime factorization $n = p_1^{a_1} p_2^{a_2} \dots p_r^{a_r}$ then has $r \ge 1$ and we see that $\sum_{d\,|\,n} \mu(d) = \sum_{d\,|\,p_1\dots p_r} \mu(d) = \sum_{i=0}^r \binom{r}{i}(-1)^i = (1-1)^r = 0.$

2. By the assumption, $\sum_{d\,|\,n} \mu(d) f(n/d) = \sum_{d\,|\,n} \mu(d) \sum_{e\,|\,(n/d)} g(e)$. This equals $\sum_{de\,|\,n} \mu(d) g(e) = \sum_{e\,|\,n} g(e) \sum_{d\,|\,(n/e)} \mu(d) = g(n)$ by part 1. $\square$

The *von Mangoldt function*

$$\Lambda:\ \mathbb{N} \to \mathbb{R}$$

is defined by $\Lambda(n) = \log p$ if $n = p^k$ for a prime p and $k \geq 1$ and $\Lambda(n) = 0$ else.

Proposition 2.1.7 *The von Mangoldt function Λ has the following properties.*

1. *For every $n \in \mathbb{N}$, $\sum_{d\,|\,n} \Lambda(d) = \log n$.*
2. *For every $n \in \mathbb{N}$, $\Lambda(n) = -\sum_{d\,|\,n} \mu(d) \log d$.*
3. *For every $n \in \mathbb{N}$ and $x > 0$, the sum $\sum_{d\,|\,n} \mu(d) \log(x/d)$ equals $\log x$ if $n = 1$ and $\Lambda(n)$ otherwise.*
4. *For $x > 1$, $\sum_{n<x} \Lambda(n) = O(x)$.*
5. *For $x > 1$, $\sum_{n<x} \Lambda(n)/n = \log x + O(1)$.*

Proof. 1. Considering the prime factorization $n = p_1^{a_1} p_2^{a_2} \dots p_r^{a_r}$ we see that $\sum_{d\,|\,n} \Lambda(d) = \sum_{p^k\,|\,n} \log p = \sum_{i=1}^{r} a_i \log p_i = \log n$.

2. Möbius inversion of part 1 and part 1 of Proposition 2.1.6 give $\Lambda(n) = \sum_{d\,|\,n} \mu(d) \log(n/d) = (\log n) \sum_{d\,|\,n} \mu(d) - \sum_{d\,|\,n} \mu(d) \log d = -\sum_{d\,|\,n} \mu(d) \log d$.

3. We write $\sum_{d\,|\,n} \mu(d) \log(x/d) = (\log x) \sum_{d\,|\,n} \mu(d) - \sum_{d\,|\,n} \mu(d) \log d$ and use part 1 of Proposition 2.1.6 and the previous result.

4. The sum equals $\sum_{p^k<x} \log p$. The contribution of prime powers with $k \geq 2$ is small, at most $x^{1/2}(\log x/\log 2)\log x = O(x^{1/2}(\log x)^2)$, and it suffices to show that $\sum_{p<x} \log p = O(x)$. Now $\prod_{n<p\leq 2n} p \leq \binom{2n}{n} < 4^n$ for every $n \in \mathbb{N}$ because every prime in the range divides $\binom{2n}{n} = 2n(2n-1)\dots(n+1)/n!$ and $\binom{2n}{n} < (1+1)^{2n}$ by the binomial expansion. Hence $\sum_{n<p\leq 2n} \log p < (\log 4)n$. Summing these inequalities with $n = 1, 2, 4, 8, \dots, 2^k$ where $2^k < x \leq 2^{k+1}$ we get that $\sum_{p<x} \log p < (\log 4)(1 + 2 + 4 + 8 + \dots + 2^k) < 2(\log 4)x$.

5. We sum the identities of part 1 over $n < x$, interchange summations and get $\sum_{d<x} \Lambda(d)\lfloor x/d \rfloor = \sum_{n<x} \log n$. Integral estimate shows that the right side is $\int_1^x \log(t)\; dt + O(\log x) = x \log x + O(x)$. By part 4, the left side equals $x \sum_{d<x} \Lambda(d)/d + O(x)$. Dividing by x we obtain the stated asymptotics. $\square$

2.2 Proof of Dirichlet's theorem

For χ a Dirichlet character modulo m we consider the infinite series

$$L(1, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n}.$$

If $\chi = \chi_0$, it diverges to $+\infty$ because it is minorized by the divergent series $1 + \frac{1}{m+1} + \frac{1}{2m+1} + \dots$. For $\chi \neq \chi_0$ it converges to a finite sum as Abel's lemma and part 2 of Proposition 2.1.2 show that $|\sum_{n=k}^{l} \chi(n)/n| \leq (\varphi(m) - 1)/k \to 0$ for $k \to \infty$. The main tool is the finite sum ($x > 0$)

$$S(x, \chi) = \sum_{n<x} \frac{\chi(n)\Lambda(n)}{n}.$$

Proposition 2.2.1 *Let χ be a Dirichlet character modulo m. For $x \to +\infty$ the following holds.*

1. *If $\chi = \chi_0$ then $S(x, \chi_0) = \log x + O(1)$.*
2. *If $\chi \neq \chi_0$ and $L(1, \chi) = 0$ then $S(x, \chi) = -\log x + O(1)$.*
3. *If $\chi \neq \chi_0$ and $L(1, \chi) \neq 0$ then $S(x, \chi) = O(1)$.*

Proof. 1. Clearly, $S(x, \chi_0) = \sum_{n<x} \Lambda(n)/n - \sum_{p \mid m,\, p^k<x} (\log p)/p^k$. The first sum on the right is $\log x + O(1)$ by part 5 of Proposition 2.1.7 and the second is at most $\sum_{p \mid m} (\log p)/(p-1) = O(1)$. Thus $S(x, \chi_0) = \log x + O(1)$.

2. By part 3 of Proposition 2.1.7, $\log x + \sum_{n<x} \chi(n)\Lambda(n)/n$ equals to the sum $\sum_{n<x} (\chi(n)/n) \sum_{d \mid n} \mu(d) \log(x/d)$. Exchanging summations, writing $n = de$ and using complete multiplicativity of χ we transform it in the double sum $D = \sum_{d<x} \mu(d) \log(x/d)(\chi(d)/d) \sum_{e<x/d} \chi(e)/e$. The inner sum equals, by the assumption, $L(1, \chi) - \sum_{e \geq x/d} \chi(e)/e = -\sum_{e \geq x/d} \chi(e)/e$.

By Abel's lemma and part 2 of Proposition 2.1.2, modulus of the last infinite sum is smaller than $\varphi(m)d/x$. Hence $|D| < (\varphi(m)/x)\sum_{d<x}\log(x/d) = O(1)$ because $\sum_{d<x}\log(x/d) = x\log x - \sum_{d<x}\log d + O(\log x)$ and the last sum equals $x\log x + O(x)$ by the integral estimate. We conclude that $\sum_{n<x}\chi(n)\Lambda(n)/n = -\log x + D = -\log x + O(1)$.

3. By part 1 of Proposition 2.1.7, $\sum_{n<x}\chi(n)(\log n)/n$ equals to the sum $\sum_{n<x}(\chi(n)/n)\sum_{d\,|\,n}\Lambda(d)$. Exchanging summations, writing $n = de$ and using complete multiplicativity of χ we transform it in the double sum $D = \sum_{d<x}(\chi(d)\Lambda(d)/d)\sum_{e<x/d}\chi(e)/e$. As in the previous part, by Abel's lemma the inner sum is $L(1,\chi)+cd/x$ where $|c| < \varphi(m)$. Hence we get that $|D - L(1,\chi)\sum_{d<x}\chi(d)\Lambda(d)/d| < (\varphi(m)/x)\sum_{d<x}\Lambda(d) = O(1)$, by part 4 of Proposition 2.1.7. Thus $\sum_{n<x}\chi(n)(\log n)/n = L(1,\chi)\sum_{d<x}\chi(d)\Lambda(d)/d + O(1)$. By Abel's lemma and part 2 of Proposition 2.1.2, modulus of the sum on the left is $O(1)$. Dividing by $L(1,\chi)$ we conclude that $\sum_{d<x}\chi(d)\Lambda(d)/d = O(1)$. □

Next we prove that the second possibility does not occur.

Proposition 2.2.2 *$L(1,\chi) \neq 0$ for every non-principal Dirichlet character χ modulo m.*

Proof. Let V be the number of non-principal Dirichlet characters χ modulo m for which $L(1,\chi) = 0$. By part 5 of Proposition 2.1.2 and Proposition 2.2.1 we have $0 \le \varphi(m)\sum_{n\equiv 1\,(m),\,n<x}\Lambda(n)/n =$
$\sum_{\chi}\sum_{n<x}\chi(n)\Lambda(n)/n = (1-V)\log x + O(1)$. So $V \ge 2$ gives a contradiction for large x and there is at most one non-principal χ with $L(1,\chi) = 0$. This χ has only real values because else $\overline{\chi}$ would be a different character and $L(1,\chi) = 0$, $L(1,\overline{\chi}) = \overline{L(1,\chi)} = 0$ would give $V \ge 2$. We assume that χ is a non-principal real character and show that $L(1,\chi) \neq 0$.

We denote $S(t) = \sum_{i\le t}\chi(i)$. By Lemma 2.1.4, $\sum_{i\le x}\chi(i)/i = S(x)x^{-1} + \int_1^x S(t)t^{-2}\,dt$. The left side differs from $L(1,\chi)$ by $O(x^{-1})$ (by Abel's lemma) and $S(x) = O(1)$ (part 2 of Proposition 2.1.2). Hence multiplying the last equality by x we get $xL(1,\chi) = O(1) + \int_1^x(\sum_{i\le t}\chi(i))x\,dt/t^2$. This integral equals, up to the error $O(\log x)$, $\int_1^x(\sum_{i\le t}\chi(i))\lfloor x/t\rfloor\,dt/t$, which we write as $\int_1^x(\sum_{i\le t,\,j\le x/t}\chi(i))\,dt/t$. By Lemma 2.1.5, it equals to $\sum_{ij\le x}\chi(i)\int_i^{x/j}dt/t = \sum_{ij\le x}\chi(i)\log(x/ij)$. Writing $n = ij$ and exchanging summations we get the sum $\sum_{n\le x}\log(x/n)\sum_{i\,|\,n}\chi(i)$. By part 3 of Proposition 2.1.2, the inner sum is always nonnegative and at least 1 for squares. Thus we get the inequality

$$xL(1,\chi) > O(\log x) + \sum_{k^2 \le x} \log(x/k^2) > O(\log x) + (\log 4)\lfloor \sqrt{x}/2 \rfloor$$

(considering k with $k \le \sqrt{x}/2$). If $L(1,\chi) = 0$ this produces contradiction for large x. Hence $L(1,\chi) \neq 0$, in fact $L(1,\chi) > 0$. $\square$

Now we prove Dirichlet's theorem. Let $a \in \mathbb{N}$ be coprime with $m \in \mathbb{N}$ and let

$$S(x,a) = \sum_{n<x,\, n \equiv a\ (m)} \frac{\Lambda(n)}{n}.$$

On the one hand, $S(x,a)$ equals $\sum_{p<x,\, p\equiv a\ (m)} (\log p)/p$ plus the contribution from the prime powers $p^k < x$, $p^k \equiv a\ (m)$, with $k \ge 2$. But this is at most $\sum_{k=2}^{\infty}\sum_{n=2}^{\infty}(\log n)/n^k = O(1)$. On the other hand, after exchanging summations and using part 5 of Proposition 2.1.2, $\varphi(m)S(x,a) = \sum_\chi \overline{\chi}(a)S(x,\chi)$, summed over all $\varphi(m)$ Dirichlet characters modulo m. By Propositions 2.2.1 and 2.2.2, the last sum is $\log x + O(1)$ because all summands with $\chi \neq \chi_0$ are $O(1)$ and the remaining summand with $\chi = \chi_0$ is $\log x + O(1)$. We conclude that, for $x > 0$,

$$\sum_{p<x,\, p\equiv a\ (m)} \frac{\log p}{p} = \frac{\log x}{\varphi(m)} + O(1).$$

This goes to infinity with x and thus there are infinitely many primes congruent to a modulo m.

2.3 Erdős's partial proof of Dirichlet's theorem

Let $a, m \in \mathbb{N}$, $1 \le a < m$ and $(a,m) = 1$, $p_1, p_2, \ldots, p_h$ be the list of all primes that are smaller than m and do not divide it and $q_i \in \mathbb{N}$, $1 \le q_i < m$, be given by $p_i q_i \equiv a$ modulo m. For $n \in \mathbb{N}$ we consider the numbers

$$\begin{aligned} P_n(a,m) &= \frac{(a+m)(a+2m)\ldots(a+nm)}{n!} \\ Q_n(a,m) &= \frac{P_n(a,m)}{P_{n/p_1}(q_1,m)P_{n/p_2}(q_2,m)\ldots P_{n/p_h}(q_h,m)}. \end{aligned}$$

In $Q_n(a,m)$ we assume for simplicity that n is divisible by $p_1p_2\dots p_h$. Both $P_n(a,m)$ and $Q_n(a,m)$ are in general fractions. For a prime p and $\frac{a}{b}\in\mathbb{Q}$ we define $\mathrm{ord}_p(a/b)\in\mathbb{Z}$, *the order of p in $\frac{a}{b}$*, as $i-j$ where p^i (resp. p^j) is the highest power of p dividing a (resp. b). We have the prime factorization $\frac{a}{b}=\prod_p p^{\mathrm{ord}_p(a/b)}$. If $\mathrm{ord}_p(a/b)\ge k$, we say that p^k divides $\frac{a}{b}$.

Proposition 2.3.1 *Let $\sigma=\sigma(m)=\sum_{i=1}^h 1/p_i$. The numbers $P_n(a,m)$ and $Q_n(a,m)$ have the following properties.*

1. *$Q_n(a,m)=m^{(1-\sigma)n+o(n)}$ as $n\to\infty$ (on $n\equiv 0 \pmod{p_1p_2\dots p_h}$).*
2. *Let $(p,m)=1$ and $k=\mathrm{ord}_p(P_n(a,m))$. Then $1\le p^k<(n+1)m$.*
3. *If p divides m and $\sigma\le 1$ then $\mathrm{ord}_p(Q_n(a,m))\le 0$.*
4. *Suppose that $n\ge m$ and $p>\sqrt{(n+1)m}$, p divides $P_n(a,m)$ and $p\not\equiv a$ modulo m. Then p divides $P_{n/p_i}(q_i,m)$ for some i, $1\le i\le h$.*

For m with $\sigma(m)<1$ we derive from this Dirichlet's theorem. We apply Proposition 2.3.1 to the prime factorization of $Q_n(a,m)$ for $n\ge m$ and divisible by $p_1p_2\dots p_h$. For a prime p let $k=\mathrm{ord}_p(Q_n(a,m))$. By parts 2 and 4, $k=0$ for $p\ge(n+1)m$, $k\le 1$ for $p>\sqrt{(n+1)m}$ and $k\le 0$ if $p>\sqrt{(n+1)m}$ and is not a modulo m. For $p\le\sqrt{(n+1)m}$ we always have (by parts 2 and 3) that $p^k<(n+1)m$. Hence, by part 1,

$$m^{(1-\sigma)n+o(n)}=Q_n(a,m)\le\prod_{p\le\sqrt{(n+1)m}}(n+1)m\prod_{\substack{p<(n+1)m\\ p\equiv a\ (m)}}p.$$

The first product is at most $((n+1)m)^{\sqrt{(n+1)m}}=m^{o(n)}$, $n\to\infty$. Therefore

$$\prod_{\substack{p<(n+1)m\\ p\equiv a\ (m)}}p>m^{(1-\sigma)n+o(n)},\ n\to\infty.$$

If $\sigma<1$, the right side goes to $+\infty$ and so does the product, which means that the congruence class a modulo m contains infinitely many primes.

It remains to prove Proposition 2.3.1. Key is property 4 sifting out primes not congruent a mod m. We begin with a simple but crucial lemma on arithmetic progressions.

Lemma 2.3.2 *Let $a\in\mathbb{Z}$, $d,m,n\in\mathbb{N}$, $A=\{a+m,a+2m,\dots,a+nm\}\subset\mathbb{Z}$ and $A(d)$ be the number of multiples of d in A.*

1. *If $(d,m) = 1$ then $A(d) = \lfloor n/d \rfloor$ or $\lfloor n/d \rfloor + 1$.*

2. *If $(d,m) = 1$ then $A(d) = \lfloor n/d \rfloor + 1$ if and only if there is a multiple $a + jm$ of d in A such that $1 \le j \le n - d\lfloor n/d \rfloor$.*

3. *If $a = 0$ and $m = 1$ then $A(d) = \lfloor n/d \rfloor$ for every n and d.*

4. *If k is the order of p in $\prod_{x \in A} x$ then $k = \sum_{i \ge 1} A(p^i)$. In fact, this holds for any subset $A \subset \mathbb{Z}$.*

Proof. 1. If $j, k \in \mathbb{Z}$ are noncongruent modulo d then so are $a + jm, a + km$ because $(d,m) = 1$. Thus if $I \subset \mathbb{Z}$ is an interval with $|I| \le d$ then the numbers $a + jm$, $j \in I$, are pairwise noncongruent modulo d. Thus if $|I| = d$ then $a + jm$ is a multiple of d for exactly one $j \in I$ and if $|I| < d$ then there is at most one such j. Partitioning $\{1, 2, \ldots, n\}$ into $\lfloor n/d \rfloor$ intervals with length d and one shorter residual interval we get the bound for $A(d)$.

2. This follows by the argument in part 1 if the residual interval is selected as $\{1, 2, \ldots, n - d\lfloor n/d \rfloor\}$.

3. This follows by parts 1 and 2 as no $a + jm = j \in \{1, 2, \ldots, n - d\lfloor n/d \rfloor\}$ is a multiple of d.

4. The identity follows by double counting the pairs $(i, x) \in \mathbb{N} \times A$ where p^i divides x. $\square$

Proof of Proposition 2.3.1. 1. From $jm < a + jm < (j+1)m$ we get that $m^n < P_n(a,m) < (n+1)m^n$ and $P_n(a,m) = m^{n + o(n)}$. Hence $Q_n(a,m) = m^{n - n/p_1 - \cdots - n/p_h + o(n)} = m^{(1-\sigma)n + o(n)}$.

2. We show that $k \ge 0$ and $p^k \le a + nm$. Denoting $A = \{a + m, a + 2m, \ldots, a + nm\}$, $B = \{1, 2, \ldots, n\}$ and using parts 1, 3 and 4 of Lemma 2.3.2 we get that $k = \sum_{i \ge 1}(A(p^i) - B(p^i))$ is nonnegative and at most j, where $p^j \le a + nm < p^{j+1}$, because $A(p^i) = B(p^i) = 0$ if $i > j$. Thus $k \ge 0$ and $p^k \le p^j \le a + nm < (n+1)m$.

3. If p divides m then it does not divide the numerator of any $P_n(a,m)$ and we see that $\mathrm{ord}_p(Q_n(a,m))$ equals minus the order of p in $n!/(n/p_1)! \ldots (n/p_h)!$. This fraction is an integer because $(n/p_1) + \cdots + (n/p_h) \le n$. Therefore $\mathrm{ord}_p(Q_n(a,m)) \le 0$.

4. As $p > m$, p is coprime with m. Clearly, p^2 does not divide any of the numbers $j, a + jm$ for $1 \le j \le n$. Therefore, in the notation of part 2, $1 = k = A(p) - B(p) = A(p) - \lfloor n/p \rfloor$. By part 2 of Lemma 2.3.2, if we denote $l = n - p\lfloor n/p \rfloor$, there exists j, $1 \le j \le l$, for which $a + jm$ is a

multiple of p (in particular, $l > 0$ and p does not divide n). So

$$a + jm = pb,\ b \in \mathbb{N} \text{ and } 1 \le j \le l.$$

We may assume that the $j \ge 1$ here is the least one. This implies, since $1 \le a < m$, $p > m$ and pb is a modulo m, that $1 \le b < m$ and $(b, m) = 1$. But $b \ne 1$ because p is not a modulo m. Thus $1 < b < m$ and b is divisible by some p_i (a prime smaller than m and not dividing m), $b = p_i c$ for $c \in \mathbb{N}$. Now $a = p_i q_i + tm$ for some $t \in \mathbb{Z}$. Since $p_i, q_i > 0$ and $a < m$, we see that $t \le 0$. Plugging in for b and a we get the equality

$$p_i q_i + (t + j)m = pp_i c.$$

As $(p_i, m) = 1$, we see that p_i divides $t + j$ and $t + j = p_i j'$ for $j' \in \mathbb{Z}$. Therefore

$$q_i + j'm = pc.$$

As for the size of j', we have $j' \ge 1$ because $pc > m$ but $1 \le q_i < m$. We define l' by $l' = (n/p_i) - p\lfloor \frac{n/p_i}{p} \rfloor$ (recall that p_i divides n) and it remains to show that $j' \le l'$. This by Lemma 2.3.2, especially part 2, gives that p divides also $P_{n/p_i}(q_i, m)$. Suppose for contrary that $0 \le l' < j'$. Then $l' < j' = (t + j)/p_i \le j/p_i$ and $0 \le p_i l' < j \le l \le p - 1$. After multiplying the equality defining l' by p_i we get that $p_i l' = l$ (by the unicity of residue upon dividing n by p), which is a contradiction. Therefore $1 \le j' \le l'$ and $P_{n/p_i}(q_i, m)$ is divisible by p. □

2.4 Remarks

The proof in Sections 2.1 and 2.2 is based on the nice exposition of Pollack [29, chapter 4], which in turn follows the presentation of Gel'fond and Linnik [18] ([17, section 3.2]) of Shapiro's proof [35]. Harrison [19] used it to give formalized computer-verified proof of Dirichlet's theorem. The proof of non-vanishing of $L(1, \chi)$ (Proposition 2.2.2) is due to Yanagisawa [44] (the book of Gel'fond and Linnik contains another proof, which we reproduced in [22]). In Section 2.3 we follow and streamline Erdős [11], who was more interested in proving analogues of Bertrand's postulate for arithmetic progressions. For his classical proof of Bertrand's postulate see the book of Aigner and Ziegler [1]. Improvements and extensions of his bounds were achieved by Moree [27], who determined the list of $m \in \mathbb{N}$ with $\sigma(m) < 1$. For other elementary approaches to Dirichlet's theorem and further information see Narkiewicz [28]. Results similar to those of Erdős were obtained earlier by Ricci in [30] and [31], and Erdős acknowledges this in his article. (Had Ricci found an erdősian proof of particular cases of Dirichlet's theorem before Erdős? By the reviews in Zentralblatt, Ricci's arguments invoke the Prime Number Theorem, which seems to render them non-elementary.)

Chapter 3

The Gel'fond–Schneider theorem on transcendence of α^β

Hermite's arithmetical theorems on the exponential function and their extension by Lindemann are certain of the admiration of all generations of mathematicians. Thus the task at once presents itself to penetrate further along the path here entered (...) We can also give this statement a geometrical form, as follows:

If, in an isosceles triangle, the ratio of the base angle to the angle at the vertex be algebraic but not rational, the ratio between base and side is always transcendental.

In spite of the simplicity of this statement and of its similarity to the problems solved by Hermite and Lindemann, I consider the proof of this theorem very difficult; as also the proof that

The expression α^β, for an algebraic base α and an irrational algebraic exponent β, e. g., the number $2^{\sqrt{2}}$ or $e^\pi = i^{-2i}$, always represents a transcendental or at least an irrational number.

It is certain that the solution of these and similar problems must lead us to entirely new methods and to a new insight into the nature of special irrational and transcendental numbers.

D. Hilbert [20]

This fragment of the address given by D. Hilbert (1862–1943) at the International Congress of Mathematicians in Paris in 1900, which in the printed version lists 23 famous problems, concerns the seventh problem on transcendental numbers. Hilbert regarded it as difficult and later in 1920s in a popular lecture stated that he may live to see solution of Riemann's hypothesis (the 8th problem), the youngest people in the auditory may live to see resolution of Fermat's Last Theorem, but the solution of his seventh problem lies farther in the future. As for the FLT he was perhaps right (A. Wiles proved it with the help of R. Taylor in 1994–95) but not so in the case of the other two problems. In 2010 RH remains open (despite that in the ArXiv one sees its "(dis)proof" every other month or so) but the seventh problem was solved in Hilbert's lifetime. First breakthrough was obtained by Gel'fond in 1929 [13] by proving the transcendence of $e^{\pi} = i^{-2i}$ (and other similar numbers). In 1930 R. O. Kuzmin extended his result to α^{β} with real quadratic β. Finally, in 1934 Hilbert's seventh problem was solved completely and independently by Alexander O. Gel'fond (1906–1968) [14] and Theodor Schneider (1911–1988) [34].

Theorem 3.0.1 (Gel'fond, 1934; Schneider, 1934) *If $\alpha, \beta \in \mathbb{C}$ are algebraic numbers, $\alpha \neq 0, 1$ and $\beta \notin \mathbb{Q}$, then the number α^{β} is transcendental.*

Here α^{β} denotes any of the generally infinitely many values of $\exp(\beta \log \alpha)$ with $\exp(z) = \sum_{n \geq 0} z^n/n!$ and $\log \alpha = \log|\alpha| + i(\arg(\alpha) + 2k\pi)$ for $\arg(\alpha) \in [-\pi, \pi)$ and k running through $\mathbb{Z}$—each of these values is transcendental. For example,

$$2^{\sqrt{2}} = 2.66514\ldots \quad \text{and each of} \quad (\cos(2k\pi\sqrt{2}) + i\sin(2k\pi\sqrt{2})2^{\sqrt{2}}, \ k \in \mathbb{Z},$$

is transcendental. And so is

$$\mathrm{e}^{\pi} = 23.14069\ldots$$

and each of its complex relatives, since $\exp(\pi i) = -1 \in (\mathrm{e}^{\pi})^i$ and both i and -1 are algebraic.

The theorem will be proved in Section 3.2. The proof requires some standard results from the algebraic number theory on algebraic numbers, which are somewhat more involved than what we used in Chapter 1. Since we keep our lecture notes self-contained, we rederive these results from scratch in a convenient form in Section 3.1. We only assume familiarity of the reader with linear algebra and language of commutative algebra.

3.1 Algebraic numbers and number fields

We consider only fields that are subfields of the field $\mathbb{C}$ of complex numbers. If $K \subset L \subset \mathbb{C}$ are two such fields, we write $[L : K]$ for the *degree of L over K*, the dimension of L as a vector space over K. Every field K has $\mathbb{Q}$ as a subfield and if $d = [K : \mathbb{Q}] < \infty$, we call K a *number field* and d its *degree*. If K is a field and $X \subset \mathbb{C}$, then $K[X] = \{p(\alpha_1, \ldots, \alpha_r) \mid p \in K[x_1, \ldots, x_r], \alpha_i \in X\}$, resp. $K(X) = \{p(\alpha_1, \ldots, \alpha_r)/q(\alpha_1, \ldots, \alpha_r) \mid p, q \in K[x_1, \ldots, x_r], \alpha_i \in X\}$, is the smallest subring, resp. subfield, of $\mathbb{C}$ containing $K \cup X$. Clearly, $K[X] \subset K(X)$. For finite $X = \{\alpha_1, \ldots, \alpha_r\}$ we write $K[\alpha_1, \ldots, \alpha_r]$ and $K(\alpha_1, \ldots, \alpha_r)$ instead of $K[\{\alpha_1, \ldots, \alpha_r\}]$ and $K(\{\alpha_1, \ldots, \alpha_r\})$. For K a field and $\alpha \in \mathbb{C}$ we say that α *is algebraic over K* if $p(\alpha) = 0$ for a nonzero $p \in K[x]$; the *minimum polynomial of α over K* and the *degree of α over K* are defined in the manner analogous to the case $K = \mathbb{Q}$. As in the case $K = \mathbb{Q}$, the minimum polynomial of α over K is irreducible in $K[x]$, has only simple roots and divides every $q \in K[x]$ with root α (cf. Proposition 1.1.2).

Proposition 3.1.1 *Fields and number fields have the following properties.*

1. *If $K \subset L \subset M$ are fields then $[M : K]$ is finite iff both $[M : L]$ and $[L : K]$ are finite, and in this case $[M : K] = [M : L] \cdot [L : K]$.*

2. *If $K \subset L$ are fields with $[L : K] < \infty$ then every $\alpha \in L$ is algebraic over K. In particular, every element of a number field is an algebraic number.*

3. *If $X \subset \mathbb{C}$ is a finite set of numbers algebraic over a field K then $K(X) = K[X]$ and $[K(X) : K] < \infty$. In particular, $\mathbb{Q}(X) = \mathbb{Q}[X]$ and is a number field if X is a finite set of algebraic numbers.*

4. *If $\alpha \in \mathbb{C}$ is algebraic over a field K then $K(\alpha) = K[\alpha]$ and $[K(\alpha) : K]$ equals to the degree of α over K.*

Proof. 1. Suppose that $[M : L] = m$, $[L : K] = l$ are finite and $\{v_1, \ldots, v_m\} \subset M$, resp. $\{u_1, \ldots, u_l\} \subset L$, is a linear basis of M over L, resp. of L over K. It is not hard to check that $\{v_j u_i \mid 1 \le j \le m, 1 \le i \le l\} \subset M$ is then a linear basis of M over K. Thus $[M : K] = ml$ is finite. Conversely, if $[M : K] = k$ is finite, then L as a vector subspace of M has $[L : K] = l$ with $l \le k$. Each $k + 1$ elements of M are linearly dependent over K, thus over L and we see that $[M : L] = m \le k$. By the previous direction we have that $k = lm$.

2. Let $\alpha \in L$ and $[L : K] = d$. The $d+1$ elements $1, \alpha, \alpha^2, \ldots, \alpha^d$ of L are linearly dependent over K, which means that $p(\alpha) = 0$ for a nonzero polynomial $p \in K[x]$ with degree at most d.

3. Suppose that $X = \{\alpha_1, \ldots, \alpha_r\}$ where α_i has degree d_i over K. Let $d = d_1 d_2 \ldots d_r$. Then $K[X]$ as a vector space over K has dimension at most d and is a K-linear span of the d elements $\alpha_1^{i_1} \ldots \alpha_r^{i_r}$, $0 \le i_j < d_j$, because each power α_i^e, $e \in \mathbb{N}_0$, expresses as a K-linear combination of the first d_i powers $1, \alpha_i, \ldots, \alpha_i^{d_i - 1}$ (clear for $0 \le e \le d_i$ and for $e > d_i$ we use in $\alpha_i^e = \alpha_i \alpha_i^{e-1}$ induction). Thus also $[K(X) : K] \le d$ because every $d+1$ elements $a_1, \ldots, a_{d+1}$ of $K(X)$ are linearly dependent over K; there is a nonzero $a \in K[X]$ such that $aa_i \in K[X]$ for every $1 \le i \le d+1$ and since aa_i are linearly dependent over K, so are a_i. This argument shows that in fact the dimension of $K[X]$ over K equals $[K(X) : K]$ and in $K[X] \subset K(X)$ we have equality.

4. By 3, $K(\alpha) = K[\alpha]$. Let α have degree d over K. As we know, every power α^e, $e \in \mathbb{N}_0$, is a K-linear combination of $1, \alpha, \ldots, \alpha^{d-1}$ and these are linearly independent over K (else α would have a smaller degree over K). Thus $\{1, \alpha, \ldots, \alpha^{d-1}\}$ is a linear basis of $K[\alpha]$ over K and $[K(\alpha) : K] = d$. □

If $K \subset L$ are fields with finite $[L : K] = l$, then in an obvious way we can reach L from K by a chain of one-element algebraic extensions $K = K_0 \subset K_1 \subset \ldots \subset K_r = K$ where $K_i = K_{i-1}(\alpha_i)$, $0 < i \le r$, for an α_i with degree $d_i > 1$ over K_{i-1}. By 1 and 4 of the proposition, $l = d_1 d_2 \ldots d_r$. Let us call this *a simple chain*. It can be proven that a simple chain exists with $r = 1$, in particular every number field is of the form $\mathbb{Q}(\alpha)$ for an algebraic $\alpha \in \mathbb{C}$, but we will not need this.

For an algebraic number $\alpha \in \mathbb{C}$ with degree d, the *set of conjugates of* α is the set $\mathrm{con}(\alpha)$ of all roots of the minimum polynomial of α. So $\alpha \in \mathrm{con}(\alpha)$ and $|\mathrm{con}(\alpha)| = d$. An *embedding* of a field K (in $\mathbb{C}$) is any (necessarily injective) field homomorphism $\sigma : K \to \mathbb{C}$. We denote the set of all embeddings of K by $G(K)$. So $\mathrm{id}_K \in G(K)$ and $\sigma|\mathbb{Q} = \mathrm{id}_{\mathbb{Q}}$ for every $\sigma \in G(K)$ and every field K. If $p \in K[x_1, \ldots, x_r]$, $\sigma(p)$ denotes the polynomial obtained from p by applying σ to the coefficients.

Proposition 3.1.2 *Embeddings of fields and conjugates of algebraic numbers relate as follows.*

1. *If $K \subset L$ are fields with $[L : K] = l < \infty$ then every $\sigma \in G(K)$ has exactly l extensions to an embedding of L.*

2. *Every number field L with degree d has exactly d embeddings.*

3. *Let $\alpha \in \mathbb{C}$ be algebraic of degree d and let $\alpha \in L$ for a degree e number field L. Then d divides e,*

$$\{\sigma(\alpha) \mid \sigma \in G(L)\} = \mathrm{con}(\alpha)$$

and for every $\beta \in \mathrm{con}(\alpha)$ there are exactly e/d embedding σ of L with $\sigma(\alpha) = \beta$.

Proof. 1. It suffices to prove this when $L = K(\alpha) = K[\alpha]$ where α has degree l over K; the general case follows by reaching L from K by a simple chain. Let $p \in K[x]$ be the minimum polynomial of α over K and $\sigma \in G(K)$. Each extension $\tau \in G(L)$ of σ sends α to a root of $\sigma(p)$ because $0 = \tau(p(\alpha)) = \tau(p)(\tau(\alpha)) = \sigma(p)(\tau(\alpha))$. The value $\beta = \tau(\alpha)$ uniquely determines τ because every $a \in L$ has form $a = q(\alpha)$ for some $q \in K[x]$ and so $\tau(a) = \sigma(q)(\beta)$. Fixing a root β of $\sigma(p)$ and defining $\tau = \tau_\beta : L \to \mathbb{C}$ by $q(\alpha) \mapsto \sigma(q)(\beta)$, $q \in K[x]$, we only need to show that the value $\tau(a)$ does not depend on q; if the definition is correct then clearly $\tau \in G(L)$ and extends σ. Let $a = q(\alpha) = r(\alpha)$ for $q, r \in K[x]$. Then $(q - r)(\alpha) = 0$, p divides $q - r$ and $\sigma(p)$ divides $\sigma(q - r) = \sigma(q) - \sigma(r)$, which implies that $\sigma(q)(\beta) = \sigma(r)(\beta)$. Thus both representations of a give the same value of τ. We conclude that the extensions $\tau \in G(L)$ of σ bijectively correspond to the roots of $\sigma(p)$. The polynomial $p \in K[x]$ is monic, has degree l and only simple roots. It follows that $\sigma(p)$ has the same properties and thus there are exactly l extensions τ.

2. A particular case of 1 with $K = \mathbb{Q}$, as $\mathbb{Q}$ has just the identical embedding in $\mathbb{C}$.

3. We set $K = \mathbb{Q}(\alpha)$. Considering the extensions $\mathbb{Q} \subset K \subset L$, 1 and 4 of Proposition 3.1.1 give that $d = [K : \mathbb{Q}]$ and divides $e = [L : \mathbb{Q}]$ with $e/d = [L : K]$. Every $\sigma \in G(L)$ sends α to a root β of the minimum polynomial $p \in \mathbb{Q}[x]$ of α, as $0 = \sigma(p(\alpha)) = \sigma(p)(\sigma(\alpha)) = p(\sigma(\alpha))$, and $\sigma|K \in G(K)$. We know from 1 and 2 that $\alpha \mapsto \beta$ extends to a unique $\rho \in G(K)$ which is thus a restriction to K of every $\sigma \in G(L)$ satisfying $\sigma(\alpha) = \beta$. The number of $\sigma \in G(L)$ with $\sigma(\alpha) = \beta$ therefore equals to the number of extensions of this ρ to an embedding of L. By part 1, this number is $[L : K] = e/d$. $\square$

The *size* $h(\alpha)$ of an algebraic number $\alpha \in \mathbb{C}$ is

$$h(\alpha) = \max_{\beta \in \mathrm{con}(\alpha)} |\beta|.$$

The size $h(P)$ of a polynomial P with algebraic coefficients is the maximum size of its coefficient.

Proposition 3.1.3 *Let $c \in \mathbb{Q}$ and $\alpha, \beta \in \mathbb{C}$ be algebraic numbers.*

1. *If $\alpha \in K$ for a number field K then $h(\alpha) = \max_{\sigma \in G(K)} |\sigma(\alpha)|$.*
2. *If α is a nonzero algebraic integer then $h(\alpha) \geq 1$.*
3. *It holds that $h(c) = |c|$, $h(c\alpha) = |c|h(\alpha)$, $h(\alpha + \beta) \leq h(\alpha) + h(\beta)$ and $h(\alpha\beta) \leq h(\alpha)h(\beta)$.*

Proof. 1. This follows from part 3 of Proposition 3.1.2.

2. The product of conjugates of α is the constant coefficient of the minimum polynomial of α, which is a nonzero integer. Thus in absolute value one of the conjugates must be at least 1.

3. That $h(c) = |c|$ is clear from the definition. We put $c, \alpha, \beta \in K = \mathbb{Q}(\alpha, \beta)$. Then K is a number field and $\sigma(\alpha+\beta) = \sigma(\alpha)+\sigma(\beta)$ and $\sigma(\alpha\beta) = \sigma(\alpha)\sigma(\beta)$ for every $\sigma \in G(K)$. The equality $h(c\alpha) = |c|h(\alpha)$ and the two inequalities now follow by part 1. $\square$

The set K_I of *K-integral elements* of a field K consists of all algebraic integers in K. It is easy to see that $\mathbb{Q}_I = \mathbb{Z}$ and $\mathbb{Z} \subset K_I$ for every K. Recall that each algebraic $\alpha \in K$ has a denominator, a number $k \in \mathbb{N}$ such that $k\alpha \in K_I$. It follows that any finite set $X \subset K$ of algebraic numbers has a *common denominator*, a $k \in \mathbb{N}$ such that $k\alpha \in K_I$ for every $\alpha \in X$.

Proposition 3.1.4 *K-integral elements of a field K have the following properties.*

1. *K_I is a subring of K.*
2. *For $n \in \mathbb{N}$ let $f(n) = f_K(n)$ be the number of $\alpha \in K_I$ with $h(\alpha) < n$. If K is a number field with degree d then*

$$f(n) < (2n)^{d^2}.$$

3. *If $\alpha \in K_I$ and K is a number field then $\prod_{\sigma \in G(K)} \sigma(\alpha) \in \mathbb{Z}$ and this product is nonzero iff α is nonzero.*

Proof. 1. We need to show that the set of algebraic integers is closed to sums and products. Let α, β be nonzero algebraic integers with degrees d, e and $\gamma = \alpha + \beta$. We show that $p(\gamma) = 0$ for a monic $p \in \mathbb{Z}[x]$; the proof for $\alpha\beta$ is similar. Let $c = de$ and $\{\alpha_1, \alpha_2, \ldots, \alpha_c\} = \{\alpha^i \beta^j \mid 0 \leq i < d, 0 \leq j < e\}$. As we know, every monomial $\alpha^i \beta^j$, $i, j \in \mathbb{N}_0$, is a $\mathbb{Z}$-linear combination of the elements $\alpha_1, \ldots, \alpha_c$. Hence $\gamma \alpha_i = \sum_{j=1}^{c} m_{i,j} \alpha_j$ for some $m_{i,j} \in \mathbb{Z}$ for every $1 \leq i \leq c$. We consider the $c \times c$ matrix $M = (\gamma \delta_{i,j} - m_{i,j})$. Its rows give a system of c homogeneous linear equations with c unknowns x_j, which has a nontrivial solution $x_j = \alpha_j$. Thus $\det(M) = 0$. Clearly, the determinant is a monic integral polynomial in γ with degree c, which is what we wanted.

2. We associate with each $\alpha \in K_I$ the degree d polynomial

$$q_\alpha(x) = \prod_{\sigma \in G(K)} (x - \sigma(\alpha)) = x^d + \sum_{i=1}^{d} a_i x^{d-i}.$$

By part 3 of Proposition 3.1.2, $q_\alpha(x) = p(x)^{d/e}$ where $p \in \mathbb{Z}[x]$ is the minimum polynomial of α and e is the degree of α. Thus $a_i \in \mathbb{Z}$ for $1 \leq i \leq d$. By part 1 of Proposition 3.1.3, for $h(\alpha) < n$ we have that $|a_i| = |\sum_{\sigma_1, \ldots, \sigma_i} \sigma_1(\alpha) \ldots \sigma_i(\alpha)| < \binom{d}{i} n^i$ (we sum over all i-element subsets of $G(K)$). The number of monic degree d polynomials with integral coefficients that satisfy these bounds is smaller than $\prod_{i=1}^{d} 2\binom{d}{i} n^i$. The mapping $\alpha \mapsto q_\alpha(x)$ is at most d to 1 (α is a root of q_α) and therefore

$$f(n) < d2^d \prod_{i=1}^{d} \binom{d}{i} n^i \leq (2n)^{d^2}$$

—induction shows that $1 + 2 + \cdots + d = \binom{d+1}{2} \leq d^2$ and $d2^d \prod_{i=1}^{d} \binom{d}{i} \leq 2^{d^2}$ for every $d \in \mathbb{N}$.

3. This product is up to sign the constant coefficient $a_d \in \mathbb{Z}$ of the polynomial $q_\alpha(x)$ in part 2. The last claim follows from the fact that $\sigma(\alpha) = 0$ iff $\alpha = 0$ for every $\sigma \in G(K)$ as σ is injective (and of course from the fact that $\mathbb{C}$ has no zero divisors). ☐

A well-known result, which we will not need, is the existence of *integral basis* for K_I, which is a linear basis $\{\alpha_1, \ldots, \alpha_d\} \subset K_I$ for K over $\mathbb{Q}$, $d = [K : \mathbb{Q}]$, such that $K_I = \{\sum_{i=1}^{d} a_i \alpha_i \mid a_i \in \mathbb{Z}\}$.

The next three propositions are auxiliary results for the proof of the Gel'fond–Schneider theorem in the next section. The following version of Siegel's lemma for number fields will suffice for the proof.

Proposition 3.1.5 *Let K be a number field with degree d. Any system of m homogeneous linear equations with n unknowns*

$$\sum_{j=1}^{n} a_{i,j} x_j = 0, \ 1 \le i \le m,$$

in which $n \ge 2d^2 m$ and $a_{i,j} \in K_I$ are K-integral coefficients satisfying $h(a_{i,j}) < A$, $A \ge 1$, for every i, j, has an integral solution $(\alpha_1, \dots, \alpha_n) \in \mathbb{Z}^n$ with some α_j nonzero and satisfying $|\alpha_j| < 3nA$ for every j.

Proof. Let $f_i(x_1, \dots, x_n) = \sum_{j=1}^{n} a_{i,j} x_j$, $1 \le i \le m$. For $k \in \mathbb{N}$ there are exactly k^n n-tuples in the box $B = \{0, 1, \dots, k-1\}^n$ and for every $\overline{x} \in B$ and i we have $f_i(\overline{x}) \in K_I$ and $h(f_i(\overline{x})) < knA$ (by 3 of Proposition 3.1.3). By 2 of Proposition 3.1.4 there are less than $(2knA)^{md^2}$ m-tuples in K_I^m whose components satisfy this bound. If $k^n > (2knA)^{md^2}$, two distinct n-tuples of B are mapped by the f_i to the same m-tuple and their difference $(\alpha_1, \dots, \alpha_n)$ is an integral solution to the system such that not all α_j are zero and $|\alpha_j| < k$ for every j. As $n \ge 2d^2 m$, $k > \sqrt{2knA}$ is needed, which is true for $k = \lfloor 3nA \rfloor$. □

Proposition 3.1.6 *Let K be a number field with degree d and $\alpha_1, \dots, \alpha_r \in K$. Then for every polynomial $P \in K_I[x_1, \dots, x_r]$ with K-integral coefficients and degree n,*

$$P(\alpha_1, \dots, \alpha_r) \ne 0 \Rightarrow |P(\alpha_1, \dots, \alpha_r)| > \Big(h(P)c^n\Big)^{-d}$$

where $c \ge 1$ depends only on the elements α_i.

Proof. We set $c = 2^r ka$ where $a = 1 + \max_{1 \le i \le r} h(\alpha_i) \ge 1$ and $k \in \mathbb{N}$ is a common denominator of the elements α_i. Let $P(\alpha_1, \dots, \alpha_r) \ne 0$. Since P is a sum of at most $(n+1)^r \le 2^{rn}$ monomials with coefficients of size at most $h(P)$ and degrees at most n, for every $\sigma \in G(K)$ we have $|\sigma(k^n P(\alpha_1, \dots, \alpha_r))| = k^n |\sigma(P)(\sigma(\alpha_1), \dots, \sigma(\alpha_r))| \le k^n 2^{rn} h(P) a^n = h(P)c^n$. As $k^n P(\alpha_1, \dots, \alpha_r)$ is K-integral and nonzero (1 of Prop. 3.1.4), by 3 of Proposition 3.1.4 is $\prod_{\sigma \in G(K)} \sigma(k^n P(\alpha_1, \dots, \alpha_r))$ a nonzero integer. It follows that

$$k^n |P(\alpha_1, \dots, \alpha_r))| \ge \prod_{\sigma \in G(K),\, \sigma \ne \mathrm{id}} |\sigma(k^n P(\alpha_1, \dots, \alpha_r))|^{-1} \ge (h(P)c^n)^{1-d},$$

which implies the stated bound (note that $h(P) \geq 1$). $\square$

Actually, we will need this only for polynomials $P \in \mathbb{Z}[x_1, \ldots, x_r]$ but the proof for them is not really simpler than in the present more general case.

Proposition 3.1.7 *Let $\alpha_i, \beta_i \in \mathbb{C}$, $1 \leq i \leq r$, be nonzero numbers and β_i be pairwise distinct. Then the entire function*

$$f(z) = \alpha_1 \exp(\beta_1 z) + \alpha_2 \exp(\beta_2 z) + \cdots + \alpha_r \exp(\beta_r z)$$

is not identically zero.

Proof. By induction on r. It is clearly true for $r = 1$. For $r > 1$ we consider the derivative of $f(z)\exp(-\beta_1 z)$. It equals $\sum_{i=2}^{r} \alpha_i(\beta_i - \beta_1)\exp((\beta_i - \beta_1)z)$, which is not identically zero by inductive assumption. Thus $f(z)$ is not identically zero. $\square$

3.2 Proof of the Gel'fond–Schneider theorem

Let $\alpha, \beta \in \mathbb{C}$ be algebraic, $\alpha \neq 0, 1$ and $\beta \notin \mathbb{Q}$. We fix a value of $\log \alpha$ and assume for contradiction that $\gamma = \alpha^\beta = \exp(\beta \log \alpha)$ is also algebraic. We take the number field $K = \mathbb{Q}(\alpha, \beta, \gamma)$ with degree $d = [K : \mathbb{Q}]$ and consider the (entire, as we will see) function

$$G(z) = \frac{\sum_{i,j=1}^{r} a_{i,j} \exp((\beta i + j)z)}{\prod_{t=1}^{m}(z - t \log \alpha)^s}$$

where the $a_{i,j} \in \mathbb{Z}$ and $r, m, s \in \mathbb{N}$ are appropriately chosen parameters. We show that there is a $t_0 \in \mathbb{N}$, $1 \leq t_0 \leq m$, such that $G(t_0 \log \alpha) \neq 0$. Since the numerator of $G(t_0 \log \alpha)$ lies in K and K is a number field (now we use the algebraicity of γ), we can bound $|G(t_0 \log \alpha)|$ from below using Proposition 3.1.6. On the other hand, an analytic argument bounds this number from above. We will see that for an appropriate choice of parameters these bounds contradict each other. Thus the presumed algebraicity of $\gamma = \alpha^\beta$ leads to a contradiction.

We define the parameters $a_{i,j} \in \mathbb{Z}$ and $r, m \in \mathbb{N}$. We want that the entire function

$$F(z) = \sum_{i,j=1}^{r} a_{i,j} \exp((\beta i + j)z),\ a_{i,j} \in \mathbb{Z},$$

is not identically zero, has at each point $z = t\log\alpha$, $1 \le t \le m$, zero of order at least n and at the same time $r \in \mathbb{N}$ and $|a_{i,j}|$ are not too big. We take $m \in \mathbb{N}$ as fixed, depending only on the degree d of K (at the end we will see that $m = 2d+4$ suffices for obtaining contradiction), let r run through the multiples $2d^2m^2, 2d^2(m^2+m), 2d^2(m^2+2m), \ldots$ of $2d^2m$ and set $n = r^2/2d^2m$. Then $r^2 = 2d^2mn$, $n \in \mathbb{N}$ and $n \ge rm$. As we said, we require that

$$F^{(k)}(t\log\alpha) = \sum_{i,j=1}^{r} a_{i,j}(\beta i + j)^k(\gamma^i\alpha^j)^t = 0,\ 1 \le t \le m, 0 \le k < n.$$

This gives a system of mn homogeneous linear equations with r^2 unknowns $a_{i,j}$. We make the coefficients $(\beta i+j)^k(\gamma^i\alpha^j)^t$ in the system K-integral by multiplying the equations by w^{n+2mr} where $w \in \mathbb{N}$ is a common denominator of α, β and γ. Since $r^2 = 2d^2mn$, by Proposition 3.1.5 there exist $a_{i,j} \in \mathbb{Z}$, $1 \le i, j \le r$, not all zero, for which $F(z)$ has the required zeros and which satisfy $|a_{i,j}| < 3r^2A$ where $A = \max_{i,j,k,t} h(w^{n+2mr}(\beta i+j)^k(\gamma^i\alpha^j)^t)$. By $c_1, c_2, \ldots$ we will denote positive constants depending only on α and β. Due to $n \ge rm$ we have

$$A < w^{n+2mr}(r(1+h(\beta)))^n(h(\gamma)h(\alpha))^{rm} \le (c_1r)^n$$

and see that

$$|a_{i,j}| < (4c_1r)^n,\ 1 \le i, j \le r.$$

By Proposition 3.1.7 the function $F(z)$ is not identically zero, because not all $a_{i,j}$ are zero and the $\beta i + j$ are distinct and nonzero for distinct pairs i, j due to $\beta \notin \mathbb{Q}$.

We define the parameter $s \in \mathbb{N}$. We set s to be the minimum order of a zero of $F(z)$ at some $z = t\log\alpha$, $1 \le t \le m$. By the properties of $F(z)$ is s well defined and $s \ge n$. There is a $t_0 \in \mathbb{N}$, $1 \le t_0 \le m$, such that F has at $z = t_0\log\alpha$ zero of order exactly s; at each $z = t\log\alpha$, $1 \le t \le m$, has F zero of order at least s. Now is $G(z)$ completely defined and is an entire function as the zeros of the denominator are canceled by the zeros of $F(z)$. Expanding $F(z)$ in the Taylor series centered at $t_0\log\alpha$, we get

$$G(t_0\log\alpha) = \frac{F^{(s)}(t_0\log\alpha)}{s!}\prod_{1\le t\le m,\ t\ne t_0}((t_0-t)\log\alpha)^{-s} \ne 0.$$

In the next two steps we bound $|G(t_0\log\alpha)|$ from below and from above.

We have

$$F^{(s)}(t_0 \log \alpha) = \sum_{i,j=1}^{r} a_{i,j}(\beta i + j)^s (\gamma^i \alpha^j)^{t_0} = P(\alpha, \beta, \gamma) \neq 0,\ P \in \mathbb{Z}[x, y, z].$$

P has degree at most $s + 2mr$ and coefficients $a_{i,j}\binom{s}{l} i^l j^{s-l}$, $0 \leq l \leq s$, $1 \leq i, j \leq r$, which implies (by $n \leq s$) that $h(P) = \|P\| < \max_{i,j} |a_{i,j}|(2r)^s \leq (c_2 r^2)^s$. Using Proposition 3.1.6, $r^2 = 2d^2 mn \leq 2d^2 ms$ and $mr \leq n \leq s$, we get that

$$|F^{(s)}(t_0 \log \alpha)| > \Big((c_2 r^2)^s c^{s+2mr} \Big)^{-d} > \Big(s^s (c_3 m)^s \Big)^{-d}.$$

Employing the simple estimates $1/s! \geq 1/s^s$ and $|\prod_{\dots} \dots| \geq 1/(m|\log \alpha|)^{ms}$, we get the lower bound

$$|G(t_0 \log \alpha)| > s^{s(-d-1)} (c_4 m)^{-smd}.$$

The upper bound follows by complex analysis: by the maximum modulus principle,

$$|G(t_0 \log \alpha)| \leq \max_{|z|=R} |G(z)| \leq \max_{|z|=R} |F(z)| \cdot \prod_{t=1}^{m} (R - t|\log \alpha|)^{-s}$$

for any $R > m|\log \alpha|$. We have (by $r^2 \leq 4^s$) that

$$\max_{|z|=R} |F(z)| \leq r^2 \max_{i,j} |a_{i,j}| \max_{i,j,|z|=R} |\exp((\beta i + j)z)| \leq (16 c_1 r)^s \exp(rR(1+|\beta|)).$$

Using the bound $|\prod_{\dots} \dots| < (R/2)^{-ms}$ (if $R > 2m|\log \alpha|$) and that $r \leq d\sqrt{2ms} \leq dm\sqrt{s}$, we get the upper bound

$$|G(t_0 \log \alpha)| < s^{s/2} c_5^{m\sqrt{s}R} R^{-ms} c_6^{ms}.$$

Finally, we let s go to infinity (recall that $s \geq r^2/2d^2 m$ and $r \to \infty$) and set $R = \sqrt{s}$. The upper bound then becomes

$$|G(t_0 \log \alpha)| < s^{s(1-m)/2} (c_5 c_6)^{ms}.$$

For $m = 2d + 4$ or larger and $s \to \infty$ this is smaller than the above lower bound on $|G(t_0 \log \alpha)|$, which is a contradiction.

3.3 Remarks

The proof in Section 3.2 is based on "Appendix I. The Transcendence of e and π" in Lang [23, pp. 867–873]. (In retrospect, this was not the best choice to learn about the Gel'fond–Schneider theorem but so it was.) Lang proves a more general result (due originally to him), which can be found also in Baker [4]. Further sources for the Gel'fond–Schneider theorem are Filaseta [12] and Gel'fond and Linnik [18], [17] (in both cases only for real $\alpha > 0, \beta$) and Baker and Wüstholz [5]. For the proofs of unproved results on number fields mentioned in Section 3.1 (and many more) see Marcus [25]. For stronger versions of Siegel's lemma for number fields see Bombieri and Gubler [7]. For the biografies of Siegel, Gel'fond and Schneider and information on Hilbert's problems (including the full text of Hilbert's address) see Yandell [45].

Clearly, $\exp(\beta \log \alpha) \neq \alpha'$ for algebraic α' means that $\beta \log \alpha - \log \alpha' \neq 0$. So an equivalent formulation of Theorem 3.0.1 is that if $\alpha_1, \alpha_2 \in \mathbb{C}$ are nonzero algebraic numbers whose logarithms are linearly independent over $\mathbb{Q}$ then they are linearly independent over algebraic numbers, $\beta_1 \log \alpha_1 + \beta_2 \log \alpha_2 \neq 0$ whenever $\beta_i \in \mathbb{C}$ are algebraic, not both zero. Generalization of this result to linear forms with more than two logarithms was achieved by Baker ([3] and [4]), and he opened by this a new era in number theory. But this will be topic of another text.

Bibliography

[1] M. Aigner and G. M. Ziegler, *Proofs from THE BOOK*, Springer, 2001.

[2] A. Baker, Rational approximations to $\sqrt[3]{2}$ and other algebraic numbers, *Quart. J. Math. Oxford* **15** (1964), 375–383.

[3] A. Baker, Linear forms in the logarithms of algebraic numbers I, II, III, IV, *Mathematika* **13** (1966), 204–216; **14** (1967), 102–107, 220–228; **15** (1968), 204–216.

[4] A. Baker, *Transcendental Number Theory*, Cambridge University Press, 1990 (2nd edition).

[5] A. Baker and G. Wüstholz, *Logarithmic Forms and Diophantine Geometry*, Cambridge University Press, 2007.

[6] Yu. F. Bilu and R. Tichy, The Diophantine equation $f(x) = g(y)$, *Acta Arith.* **95** (2000), 261–288.

[7] E. Bombieri and W. Gubler, *Heights in Diophantine Geometry*, Cambridge University Press, 2006.

[8] M. Davis, H. Putnam and J. Robinson, The decision problem for exponential diophantine equations, *Ann. of Math.* **74** (1961), 425–436.

[9] G. Lejeune Dirichlet, Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält, *Abh. der Königlich Preuss. Akad. der Wiss.* (1837), 45–81.

[10] F. J. Dyson, The approximation to algebraic numbers by rationals, *Acta Math.* **79** (1947), 225–240.

[11] P. Erdős, Über die Primzahlen gewisser arithmetischer Reihen, *Math. Z.* **39** (1935), 473–491.

[12] M. Filaseta, *Transcendental Number Theory*, lecture notes for MATH 785, available from the author's homepage.

[13] A. O. Gel'fond, Sur les propriétés arithmétiques des fonctions entières, *Tôhoku Math. J.* **30** (1929), 280–285.

[14] A. O. Gel'fond, On seventh problem of Hilbert, *Dokl. Akad. Nauk SSSR* **2** (1934), 1–3 (Russian).

[15] A. O. Gel'fond, *Transcendental and Algebraic Numbers*, State Publishing House of Technical and Theoretical Literature, Moscow, 1952 (Russian).

[16] A. O. Gel'fond and Ju. V. Linnik, On Thue's method in the problem of effectiveness in quadratic fields, *Doklady Akad. Nauk SSSR* **61** (1948), 773–776 (Russian).

[17] A. O. Gel'fond and Ju. V. Linnik, *Elementary Methods in Analytic Number Theory*, Fizmatgiz, Moscow, 1962 (Russian).

[18] A. O. Gel'fond and Ju. V. Linnik, *Elementary Methods in the Analytic Theory of Numbers*, Translated from the Russian by D. E. Brown. Translation edited by I. N. Sneddon. International Series of Monographs in Pure and Applied Mathematics, Vol. 92 Pergamon Press, Oxford, 1966.

[19] J. Harrison, A formalized proof of Dirichlet's theorem on primes in arithmetic progression, *J. of Formalized Reasoning* **2** (2009), 63–83.

[20] D. Hilbert, *Mathematical Problems. Lecture delivered before the International Congress of Mathematicians at Paris in 1900.* Translated from German by M. W. Newson with the author's permission for *Bull. Amer. Math. Soc.* **8** (1902), 437–479.

[21] M. Hindry and J. H. Silverman, *Diophantine Geometry. An Introduction*, Springer, 2000.

[22] M. Klazar, *Analytic and Combinatorial Number Theory I. Lecture Notes*, KAM-DIMATIA Series 2010-968, 2010.

[23] S. Lang, *Algebra*, Springer, 2002 (3rd edition).

[24] J. Liouville, Sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même reductible à des irrationnelles algébriques, *Comptes rendus de l'Académie des Science (Paris)* **18** (1844), 883–885.

[25] D. A. Marcus, *Number Fields*, Springer, 1977.

[26] Ju. V. Matijasevič, Enumerable sets are diophantine, *Soviet Math. Doklady* **11** (1970), 354–358.

[27] P. Moree, Bertrand's postulate for primes in arithmetical progressions, *Computers Math. Applic.* **26** (1993), 35–43.

[28] W. Narkiewicz, *The Development of Prime Number Theory. From Euclid to Hardy and Littlewood*, Springer, 2000.

[29] P. Pollack, *Not Always Burried Deep. A Second Course in Elementary Number Theory*, AMS, 2009.

[30] G. Ricci, Sul teorema di Dirichlet relativi alla progressione arithmetica, *Bollettino della Unione Matematica Italiana* **12** (1933), 304–309.

[31] G. Ricci, Sul teorema di Dirichlet e di Bertrand-Tchebychev relativi alla progressione arithmetica, *Bollettino della Unione Matematica Italiana* **13** (1933), 1–11.

[32] K. Roth, Rational approximation to algebraic numbers, *Mathematika* **2** (1955), 1–20. Corrigendum, *ibid.*, 168.

[33] W. M. Schmidt, *Diophantine Approximation.* Lecture Notes in Mathematics **785**, Springer, 1980.

[34] Th. Schneider, Transzendenzuntersuchungen periodischer Funktionen. I. Transzendenz von Potenzen, *J. Reine Angew. Math.* **172** (1934), 65–69.

[35] H. N. Shapiro, On primes in arithmetic progression (II), *Ann. of Math.* **52** (1950), 231–243.

[36] C. L. Siegel, Approximation algebraischer Zahlen, *Math. Zeit.* **10** (1921), 173–213.

[37] C. L. Siegel, Über einige Anwendungen diophantischer Approximationen, *Abh. Preuss. Akad. Wiss.* no. 1 (1929).

[38] V. G. Sprindžuk, *Classical Diophantine Equations in two Unknowns*, Nauka, Moskva, 1982 (Russian).

[39] J. Steuding, *Diophantine Analysis*, Chapman & Hall/CRC, 2005.

[40] A. Thue, Om en generel i store hele tal uløsbar ligning, *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl. No. 7. Kra. 1908.* (1908), 13 pp.

[41] A. Thue, Über Annäherungswerte algebraischer Zahlen, *J. Reine Angew. Math.* **135** (1909), 284–305.

[42] A. Thue, Berechnung aller Lösungen gewisser Gleichungen von der form $ax^r - by^r = f$, *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl. No. 4. Kra. 1919.* (1918), 9 pp.

[43] *Selected Mathematical Papers of Axel Thue. With an introduction by Carl Ludwig Siegel*, Edited by Trygve Nagell, Atle Selberg, Sigmund Selberg, Knut Thalberg, Universitetforlaget, Oslo-Bergen-Tromsø, 1977.

[44] N. Yanagisawa, A simple proof that $L(1,\chi) > 0$, *Sūgaku* **50** (1998), 314–315 (Japanese).

[45] B. H. Yandell, *The Honors Class. Hilbert's Problems and their Solvers*, A K Peters, 2002.

[46] U. Zannier, *Lecture Notes on Diophantine Analysis (with an appendix by Francesco Amoroso)*, Edizioni Della Normale, Pisa, 2009.