Lecture 1 : 041104

We began by identifying three (inter-connected) themes in Euclid's work :

I. Solution of Diophantine equations.

II. Multiplicatve 'structure' in $\mathbf{Z}$, in particular, the importance of *unique factorisation.*

III. The problem of 'counting' prime numbers.

Euclid's fundamental technical achievement was

Euclid's lemma *Let $a, b \in \mathbf{N}$ and let $d := SGD(a, b)$. Then d is the least positive integer which can be expressed in the form*

$$d = ax + by, \quad x, y \in \mathbf{Z}. \tag{1}$$

In particular, whenever $c \mid a$ and $c \mid b$, then $c \mid d$.

Observe that, implicit in this description of SGD, is the fact that the natural numbers are *well-ordered*, i.e.: that every non-empty subset of $\mathbf{N}$ has a smallest element. Recall also that *Euclid's algorithm* gives a general method for finding a solution to (1), which even today is the most efficient known algorithm for this problem.

From Euclid's lemma we can easily deduce the general solution to linear Diophantine equations (see **I** above), namely :

Theorem 1 *Let a, b, c be integers. There exists a solution to the linear Diophantine equation*

$$ax + by = c \tag{2}$$

if and only if $d \mid c$, where $d := SGD(a, b)$. Moreover, if in this case (x_0, y_0) is any solution to (2), then the general solution is given by

$$x = x_0 + \left(\frac{b}{d}\right) n,$$
$$y = y_0 - \left(\frac{a}{d}\right) n,$$

where $n \in \mathbf{Z}$.

Another direction leads to the development of theme **II**. Euclid's Lemma quickly gives the following fundamental property of prime numbers :

Proposition 1 *Let p be a prime and let a, b be any integers. If $p \mid ab$, then $p \mid a$ or $p \mid b$.*

This, in turn, is the most important ingredient needed to prove

Fundamental Theorem of Arithmetic *Every positive integer can be uniquely expressed as a product of primes.*

Perhaps Euclid's best-known number-theoretic proof illustrates theme **III**.

Theorem 2 *There are infinitely many primes.*

PROOF : Suppose the contrary and let $p_1, p_2, ..., p_k$ be a complete list of the primes. Consider the number

$$N := \left(\prod_{j=1}^{k} p_j \right) + 1.$$

Clearly, N is not divisible by any p_j, hence every prime factor of N is a prime not on our list - contradiction.

The three themes we have identified above greatly influenced the further development of number theory in post-Renaissance Europe, after a long interval of nearly 2000 years during which very little progress was made. Even today, these themes are central in guiding continuing research. Most of this course will be taken up with describing the early contributions of mathematicians like Fermat, Euler, Gauss, Dirichlet etc. to this process. Whenever possible, we will try to indicate more current research trends and open problems.

We begin by making some improvements to Euclid's obervation that there are infinitely many primes. Euclid's method is very easily adapted to prove

Proposition 2 *There are infinitely many primes p such that $p \equiv 3$ (mod 4).*

PROOF : Suppose the contrary and let $p_0 = 3, p_1, ..., p_k$ be the complete list of such primes. The number

$$N := 4\left(\prod_{j=1}^{k} p_j\right) + 3$$

is congruent to 3 (mod 4), hence at least one of its' prime factors also has this property. But N is clearly not divisible by any p_j - contradiction !

Considerably more effort is needed to prove

Theorem 3 *There are infinitely many primes p such that $p \equiv 1$ (mod 4).*

PROOF : We give a rather algebraic proof. The main ingredient is the following

Lemma 1 *Let p be a prime. Then there exists an integer x satisfying*

$$x^2 \equiv -1 \ (mod\ p) \tag{3}$$

if and only if $p \equiv 1$ (mod 4).

PROOF OF LEMMA 1 : We consider $\mathbf{F}_p^\times$, the multiplicative group of the finite field with p elements. It is well-known that this group is cyclic (see, for example, [3], Chapter 7, for a proof). Hence it is a cyclic group of order $p-1$. A solution to (3) corresponds to an element of order 4 in this group. There exists such an element if and only if $4 \mid p-1$, i.e.: if and only if $p \equiv 1$ (mod 4). This completes the proof of the lemma.

Now we are in a position to prove our theorem by modifying Euclid's method. Suppose, on the contrary, that there are only finitely many primes congruent to 1 (mod 4) and let $p_1, ..., p_k$ be a complete list. Consider the number

$$N := \left(2 \cdot \prod_{j=1}^{k} p_j\right)^2 + 1.$$

By the lemma, every prime factor of N must be congruent to 1 (mod 4). But N is clearly not divisible by any p_j - contradiction !

One of the highlights of the course will be a proof of the following famous result :

Theorem (Dirichlet 1829) *Let a, b be integers. A necessary and sufficient condition for there to exist infinitely mnay primes $p \equiv a \pmod{b}$ is that $SGD(a, b) = 1$.*

Note that it is clear that the condition is necessary : the hard part is to show its' sufficiency. Dirichlet's Theorem was the first big result proven using analytical methods in number theory, hence its' continuing importance. It was thus a major stepping stone to the proof of the most famous number-theoretic result of the 19th century, namely

Prime Number Theorem *For any $x > 0$, let $\pi(x)$ denote the number of primes less than or equal to x. Then*

$$\pi(x) \sim \frac{x}{\log x}.$$

A proof of this theorem is, unfortunately, beyond the scope of this course. For a detailed exposition which is close to the historical narrative, see [1].

Remark 1 For proofs of several more special cases of Dirichlet's Theorem using 'elementary' methods, the interested reader is referred to [4].

Remark 2 The following, very recently proved result, sounds superficially similar to Dirichlet's Theorem, but actually involves quite different mathematical techniques :

Theorem [2] *The set of primes contains arbitrarily long arithmetic progressions.*

This problem was open for a long time in a field called *combinatorial number theory.* The problem has a long history, dating back at least as far as a theorem of Van der Waerden from the 1920s which states that, if the positive integers are partitioned into two subsets, then at least one of them contains arbitrarily long AP:s. Proofs of various refinements, culminating in the above result, have involved methods from such diverse fields as combinatorics, harmonic analysis and ergodic theory.

References

[1] H. Davenport, Multiplicative Number Theory, Springer.
[2] B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions. Paper appeared at http://xxx.arxiv.org on April 8, 2004.
[3] I.N. Herstein, Topics in Algebra, Wiley.
[4] W. Sierpinski, Elemntary Theory of Numbers.

Lecture 2 : 081104

One theme of the work of Fermat was his efforts to find effective means of 'constructing' prime numbers. Here one is searching for a function $f : \mathbf{N} \to \mathcal{P}$, where $\mathcal{P}$ is the set of primes, which is as easy to describe as possible. Fermat had several well-known failed attempts. One example is the polynomial function

$$f(n) = n^2 + n + 41.$$

It can be checked that $f(n)$ is a prime for $1 \leq n \leq 40$. However, since $f(41)$ is clearly divisible by 41, it is not prime. In fact, it is not hard to prove

Proposition 3 *Let $f(x) \in \mathbf{Z}[x]$ be any non-constant polynomial. Then there exist infinitely many $n \in \mathbf{Z}$ such that $f(n)$ is not prime.*

PROOF : Let p be any prime dividing $f(1)$. Then for each $k > 0$, p also divides $f(p^k + 1)$. But $f(x) \to \infty$ as $x \to \infty$ for any non-constant polynomial f, hence there are infinitely many distinct composite numbers in the sequence $(f(p^k + 1))_{k=1}^{\infty}$.

Note that, on the other hand, Dirichlet's theorem says that if $f(x) = ax + b$ is any linear function s.t. $\mathrm{GCD}(a, b) = 1$, then there are infinitely many n for which $f(n)$ IS prime. The following is a famous open problem :

Open Problem *Does there exist a polynomial $f(x) \in \mathbf{Z}[x]$ of degree greater than one such that $f(n)$ is a prime for infinitely many n ?*

Another function studied by Fermat was the function $f(x) = 2^x + 1$.

Proposition 4 *Let n be a positive integer. If $2^n + 1$ is prime, then n is a power of two.*

PROOF : If a is any odd positive integer, then $x + 1$ is a factor of $x^a + 1$. Hence if we write $n = 2^k \cdot l$, where l is odd, then $2^{2^k} + 1$ is a factor of $2^n + 1$, hence the latter cannot be prime if $l > 1$.

A prime of the form $2^{2^k} + 1$ is called a *Fermat prime*. It is unknown whether there are infinitely many Fermat primes - in fact, I believe it is also unknown if the number $2^{2^k} + 1$ are composite for infinitely many k.

A similar notion is introduced by the following observation

Proposition 5 *Let a, n be positive integers with $n > 1$. If $a^n - 1$ is prime, then $a = 2$ and n is prime.*

PROOF : $x - 1$ is a factor of $x^n - 1$ for every $n > 0$. Hence $a - 1$ is a factor of $a^n - 1$, so is a proper prime factor unless $a = 2$. And if n is not prime, say $n = r \cdot s$, with $1 < r, s < n$, then $2^r - 1$ is a proper factor of $2^n - 1$, hence the latter is not prime.

A prime of the form $2^p - 1$, where p is prime, is called a *Mersenne prime*. It is also unknown whether there are infinitely many Mersenne primes, or whether $2^p - 1$ is composite for infinitely many primes p. However, since there is a far greater density of numbers of the form $2^p - 1$ than of the form $2^{2^k} + 1$, it is much easier to make numerical investigations. These have lead to explicit conjectures for the number of Mersenne primes up to x, for large x. Please look on the internet for references if you're interested.

Mersenne primes are connected to so-called *perfect numbers*, much studied by weirdos like Pythagoras and his followers in their attempts to develop a world-view based on numerology.

NOTATION : Let n be a positive integer. We denote by $d(n)$ the number of positive divisors of n, i.e.:

$$d(n) := \sum_{d|n} 1.$$

We also denote

$$\sigma(n) := \sum_{d|n} d.$$

DEFINITION : A function $f : \mathbf{N} \to \mathbf{C}$ is said to be *multiplicative* if

$$f(mn) = f(m)f(n), \qquad \text{whenever } \mathrm{GCD}(m, n) = 1.$$

Proposition 6 *(i) The functions $d(n)$ and $\sigma(n)$ are multiplicative.*
(ii) Let $f(n)$ be a multiplicative function. Set

$$g(n) := \sum_{d|n} f(d).$$

Then $g(n)$ is also multiplicative.

PROOF : Easy consequences of FTA.

DEFINITION : The *Möbius function* $\mu : \mathbf{N} \to \mathbf{N}$ is defined as follows :

$$\mu(n) := \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{if } p^2 \mid n \text{ for some prime } p, \\ (-1)^k, & \text{if } n = \prod_{i=1}^k p_i, \text{ where all } p_i \text{ are distinct.} \end{cases}$$

From the definition, it is immediate that μ is a multiplicative function. Set

$$\nu(n) := \sum_{d|n} \mu(d).$$

By Proposition 6(ii), ν is also multplicative. Let p^α be a prime power. Then

$$\nu(p^\alpha) = \sum_{i=0}^{\alpha} \mu(p^i) = \begin{cases} 1, & \text{if } \alpha = 0, \\ 0, & \text{if } \alpha > 0. \end{cases}$$

Hence, multiplicativity of ν implies that

$$\nu(n) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{if } n > 1. \end{cases}$$

This will be helpful in proving

Möbius inversion formula *Let $f, g : \mathbf{N} \to \mathbf{C}$ be any two functions. Then the following are equivalent :*

(i) for all $n \in \mathbf{N}$,

$$g(n) = \sum_{d|n} f(d),$$

(ii) for all $n \in \mathbf{N}$,

$$f(n) = \sum_{d|n} \mu(d) g(n/d).$$

PROOF : We prove that (i) $\Rightarrow$ (ii). The proof that (ii) $\Rightarrow$ (i) is similar. Since (i) holds we have that

$$\sum_{d|n} \mu(d) g(n/d) = \sum_{d|n} \mu(d) \sum_{e|\frac{n}{d}} f(e).$$

If we now change the order of summation, this double sum becomes

$$\begin{aligned}
&\sum_{e|n} f(e) \sum_{d|\frac{n}{e}} \mu(d) \\
&= \sum_{e|n} f(e)\nu(n/e) \\
&= f(n), \quad \text{v.s.v..}
\end{aligned}$$

Example : Let $\phi(n)$ denote Euler's phi-function, as usual. The Inclusion-Exclusion principle, for example, can be used to verify that

$$\phi(n) = \sum_{d|n} \mu(d) \cdot \frac{n}{d} = n \cdot \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Hence, the Möbius inversion formula (with $f(n) = \phi(n)$, $g(n) = n$) states that

$$n = \sum_{d|n} \phi(d).$$

One way to understand this equation is as follows. Let G be a (multiplicative) group of order n with generator g. Then g^k is also a generator if and only if $\mathrm{SGD}(k, n) = 1$. In other words, G has $\phi(n)$ elements of order n. For each divisor d of n, G possesses a unique cyclic subgroup of order d, which by the same reasoning contains $\phi(d)$ elements of order d.

In other words, for each divisor d of n, G contains exactly $\phi(d)$ elements of order d. Adding up, we get the above equation.

Definition : A positive integer n is called *perfect* if $\sigma(n) = 2n$.

The following problem is likely to be very hard indeed :

Open problem *Do there exist any odd perfect numbers ?*

On the other hand, even perfect numbers have been classified :

Theorem 4 (Pythagoras, Euclid, Euler) *Let n be an even number. Then n is perfect if and only if $n = 2^{p-1}(2^p - 1)$ for some prime p such that $2^p - 1$ is a (Mersenne) prime.*

Remark : Hence, it is also an open problem as to whether or not there exist infinitely many even perfect numbers.

Proof of Theorem 4 : $\Leftarrow$ Suppose $n = 2^{p-1}(2^p - 1)$, where p and $2^p - 1$ are primes. Then the divisors of n are

$$2^i, \quad i = 0, 1, ..., p-1, \qquad 2^i(2^p - 1), \quad i = 0, 1, ..., p-1.$$

Hence

$$\sigma(n) = [1 + (2^p - 1)] \left[\sum_{i=0}^{p-1} 2^i \right] = 2^p(2^p - 1) = 2n, \ v.s.v.$$

$\Rightarrow$ Let n be an even perfect number. Write $n = 2^{k-1}m$, where m is odd. By Proposition 6, $\sigma(n) = \sigma(2^{k-1})\sigma(m) = (2^k - 1)m$. Since n is perfect, we thus have

$$2^k m = (2^k - 1)\sigma(m). \tag{4}$$

Thus $2^k - 1$ divides m, say $m = M(2^k - 1)$. Substituting into (4) gives $\sigma(m) = 2^k M$. But both m and M are divisors of m, hence

$$2^k M = \sigma(m) \geq m + M = (2^k - 1)M + M = 2^k M.$$

Hence we have equality throughout, i.e.: $\sigma(m) = m + M$. In other words, m has only two divisors, which implies that m is prime and $M = 1$, thus $m = 2^k - 1$ is a Mersenne prime, v.s.v.

Lecture 3 : 081104

Non-linear Diophantine equations were already studied by the Greeks, and this theme was enthusiastically revived by Fermat and his contemporaries in the 17th century. The latter had a number of famous results of the form : 'such and such Diophantine equation has only the following solutions (maybe no solutions)'. The methods employed were basically elementary, the key often being some clever application of FTA. In particular, the following consequence of FTA was used widely :

Fact 1 *Let a, b be positive integers such that $GCD(a, b) = 1$. If ab is a k:th power, then each of a and b is itself a k:th power.*

Sometimes, Fermat and Co. got carried away in their usage of the unique factorisation idea central to FTA - we give an example below. Efforts by later generations to give rigorous proofs of their results laid the groundwork for the development in the 19th century of the body of knowledge nowadays known as *algebraic number theory*.

We start with a result which was perhaps already known to Pythagoras (and maybe even earlier civilisations, though not being an expert on these historical matters, I refrain from giving any opinion).

Theorem 5 *Let x, y, z be positive integers such that $SGD(x, y, z) = 1$ and y is odd. Then the following two statements are equivalent :*

(i) x is even and $x^2 + y^2 = z^2$,
(ii) there exist positive integers $a < b$ such that $SGD(a, b) = 1$ and

$$x = 2ab, \quad y = b^2 - a^2, \quad z = b^2 + a^2.$$

First Proof (following Baker, p.85) : Suppose (ii) holds. Then one checks directly that $x^2 + y^2 = z^2$. Let $d = \mathrm{GCD}(x, y, z)$. Then $d | b^2 \pm a^2$, hence $d | 2a^2$ and $d | 2b^2$. Thus d also divides $\mathrm{GCD}(2a^2, 2b^2) = 2$. Hence $d = 1$ or 2. But d cannot be 2, since y is odd.

Now suppose (i) holds. Write the equation as $(z + y)(z - y) = x^2$. Since both y and z are odd and $\mathrm{GCD}(y, z) = 1$, we easily deduce that

$\mathrm{GCD}(z+y, z-y) = 2$. Hence we can write

$$\left(\frac{z+y}{2}\right)\left(\frac{z-y}{2}\right) = \left(\frac{x}{2}\right)^2$$

and, by Fact 1 above, each of $\frac{1}{2}(z \pm y)$ is a perfect square, i.e.: there exist integers $a < b$ such that

$$\frac{z-y}{2} = a^2, \qquad \frac{z+y}{2} = b^2.$$

Then (ii) follows easily.

Second Proof : The proof that (ii) $\Rightarrow$ (i) is the same as before. Now suppose (i) holds. Write the equation as

$$\left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2 = 1.$$

There exists $\theta \in (0, 2\pi)$ such that $x/z = \sin\theta$, $y/z = \cos\theta$. Let $t := \tan\theta/2$. Thus

$$\frac{x}{z} = \frac{2t}{1+t^2}, \qquad \frac{y}{z} = \frac{1-t^2}{1+t^2}. \tag{5}$$

Let $r := x/z$ and $s := y/z$. Then $r^2 + s^2 = 1$. But

$$s = \frac{1-t^2}{1+t^2} \Rightarrow t = \sqrt{\frac{1-s}{1+s}} = \sqrt{\frac{1-s^2}{(1+s)^2}} = \sqrt{\frac{r^2}{(1+s)^2}} = \pm\frac{r}{1+s}.$$

The point is that $t \in \mathbf{Q}$. Let $t = a/b$ in lowest terms. Substituting into (5), we obtain

$$(b^2 + a^2)x = (2ab)z, \tag{6}$$
$$(b^2 + a^2)y = (b^2 - a^2)z. \tag{7}$$

Consider (7). Since $\mathrm{GCD}(y, z) = 1$, we must have that $y | b^2 - a^2$ and $z | b^2 + a^2$, hence there exists $\lambda \in \mathbf{N}$ such that

$$b^2 - a^2 = \lambda y, \qquad b^2 + a^2 = \lambda z. \tag{8}$$

Then λ divides $b^2 \pm a^2$, hence divides both $2a^2$ and $2b^2$, hence also divides $\mathrm{GCD}(2a^2, 2b^2) = 2$. But if $\lambda = 2$, then (8) implies that $b^2 - a^2$ is even, hence that both a and b are odd, since they can't both be even. But then $b^2 - a^2$ is divisible by 4, hence (8) implies that y is even, a contradiction.

Thus $\lambda = 1$ and it follows that (ii) holds.

A triple of relatively prime integers satisfying the equivalent conditions of Theorem 5 is called a *primitive Pythagorean triple.*

Fermat famously studied the equation $x^n + y^n = z^n$ and claimed to have proven that there are no non-trivial integer solutions for any $n > 2$. The only case known to have been proven by Fermat is the case $n = 4$. It involves a very clever application of Fact 1 above and an induction argument (which is usually referred to in this context as the *method of infinite descent*).

Theorem 6 (Fermat) *Let x, y, z be integers such that $x^4 + y^4 = z^4$. Then $xyz = 0$.*

PROOF : We consider more generally the equation

$$x^4 + y^4 = z^2 \tag{9}$$

and show that it has no integer solutions such that $xyz \neq 0$. The proof is by contradiction and makes use of Fermat's technique of *infinite descent.* More precisely, the idea is as follows : we suppose there exists a solution (x, y, z) of (9) for which $xyz \neq 0$. Then there must be a solution for which, in addition, $\mathrm{SGD}(x, y, z) = 1$. Given any such 'primitive' solution, we then show how one may construct a new one (X, Y, Z) such that $|Z| < |z|$. But since there must be, among all primitive solutions, one in which $|z|$ is minimised, we thereby obtain the desired contradiction.

So let (x, y, z) be a primitive solution to (9). Then (x^2, y^2, z) is a primitive Pythagorean triple so, by Theorem 6, if we assume WLOG that x is even and y odd, then there exist relatively prime integers a, b such that

$$x^2 = 2ab, \tag{10}$$

$$y^2 = b^2 - a^2, \tag{11}$$

$$z = b^2 + a^2. \tag{12}$$

We can rewrite (11) as

$$y^2 + a^2 = b^2,$$

and so (y, a, b) is also a primitive Pythagorean triple. Since y is odd, so must a be even, and there exist relatively prime integers p, q such that

$$a = 2pq, \tag{13}$$
$$y = q^2 - p^2, \tag{14}$$
$$b = q^2 + p^2. \tag{15}$$

Substituting (13) and (15) into (10) yields

$$x^2 = 4pq(p^2 + q^2). \tag{16}$$

Now p and q are relatively prime, hence one sees easily that both are relatively prime to $p^2 + q^2$. Thus, the three numbers p, q and $p^2 + q^2$ are pairwise relatively prime. Since their product is, by (16), a perfect square, it follows from Fact 1 that each is a perfect square. In other words, there exist pairwise relatively prime integers X, Y, Z such that

$$p = X^2, \quad q = Y^2, \quad p^2 + q^2 = Z^2.$$

Substituting the first two of these relations into the third yields the relation

$$X^4 + Y^4 = Z^2,$$

so we have constructed the desired new primitive solution to (9). It remains to check that $|Z| < |z|$. But, using (15) and (12), we have

$$Z^2 = p^2 + q^2 = b < b^2 + a^2 = z \leq z^2,$$

as required.

Lecture 4 : 101104

Theorem 7 *The only integers x, y such that*

$$y^2 + 2 = x^3 \tag{17}$$

are $x = 3, y = \pm 5$.

'PROOF' : Write (17) as

$$(y + \sqrt{-2})(y - \sqrt{-2}) = x^3. \tag{18}$$

We shall first verify that, for any integer y, the numbers $y + \sqrt{-2}$ and $y - \sqrt{-2}$ have no common factor in

$$\mathbf{Z}[\sqrt{-2}] = \{a + b\sqrt{-2} : a, b \in \mathbf{Z}\}.$$

For suppose $z := a + b\sqrt{-2}$ is a common factor. Then z divides

$$(y + \sqrt{-2}) - (y - \sqrt{-2}) = 2\sqrt{-2},$$

and, taking squares of absolute values (as complex numbers), we conclude that

$$a^2 + 2b^2 \mid 8,$$

as ordinary integers. The only possibilities are thus

(i) $a = \pm 1,\ b = 0$.
(ii) $a = \pm 2,\ b = 0$.
(iii) $a = 0,\ b = \pm 1$.
(iv) $a = 0,\ b = \pm 2$.

In case (i), $z = \pm 1$, and hence not a proper factor. In all other cases, there must exist integers c, d such that

$$y + \sqrt{-2} = (a + b\sqrt{-2})(c + d\sqrt{-2}).$$

From this it follows, by equating the real and imaginary parts, that

$$y = ac - 2bd, \tag{19}$$
$$1 = ad + bc. \tag{20}$$

Now (20) immediately rules out (ii) and (iv), since in both cases the rhs of (20) is even. But from (17) it follows already that y must be odd (otherwise the lhs of (17) will be even, but not divisible by 4), and then (19) also eliminates (iii).

Hence, we have proven that $y \pm \sqrt{-2}$ have no common factor in $\mathbf{Z}[\sqrt{-2}]$.

By (18), this implies that each of them must be a cube in $\mathbf{Z}[\sqrt{-2}]$.

Thus, there exist integers a, b such that

$$y + \sqrt{-2} = (a + b\sqrt{-2})^3. \tag{21}$$

Multiplying out the rhs of (21) and equating the real and imaginary parts yields the two equations

$$y = a^3 - 6ab^2, \tag{22}$$
$$1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2). \tag{23}$$

Immediately, (23) implies that

$$b = 3a^2 - 2b^2 = \pm 1.$$

Since a, b are integers, the only possibility is $b = 1,\ a = \pm 1$. Substituting these possibilities into (22) gives $y = \pm 5$, v.s.v.

OBS !! There is a major gap in the proof, namely the part in italics. What I state there is correct, but it requires a proof. What one would actually like to prove is a generalisation of the Fundamental Theorem of Arithmetic to the ring $\mathbf{Z}[\sqrt{-2}]$. We will return to this issue later in the course. But, for the moment, it is worth remarking that the F.T.A. does not hold in, for example, the ring $\mathbf{Z}[\sqrt{-5}]$.

Remark Eq. (17) is an example of a *Weierstraß equation* and defines a so-called *elliptic curve* over $\mathbf{Q}$. The most general form of a Weierstraß equation over $\mathbf{Q}$ (more generally over a field of characteristic other than 2 or 3) is

$$y^2 = x^3 + Ax + B, \qquad A, B \in \mathbf{Q}.$$

This defines a non-singular, so-called elliptic curve over $\overline{\mathbf{Q}}$, the algebraic closure of $\mathbf{Q}$, if and only if the *discriminant*

$$\Delta := 4A^3 + 27B^2$$

is non-zero. There is a famous theorem of Siegel that every elliptic curve over $\mathbf{Q}$ contains only finitely many integer points $(x, y) \in \mathbf{Z}^2$. For a proof, see [1], Chapter 9.

We close our informal introduction to non-linear Diophantine equations with a proof of the following attrative result, also due originally to Fermat :

Theorem 8 (Fermat 1654) *Let p be a prime. Then there exixt integers x, y such that*

$$x^2 + y^2 = p,$$

if and only if $p = 2$ or $p \equiv 1$ (mod 4).

PROOF : Om $p = 2$ så har vi lösningarna $x = \pm 1, y = \pm 1$. Om $p \equiv 3 \pmod 4$ så finns det ingen lösning, eftersom kvadraten av varje heltal är kongruent till 0 eller 1 modulo 4, så att en summa av två kvadrater är kongruent till 0,1 eller 2 modulo 4.

Antag nu att $p \equiv 1 \pmod 4$. Enligt Lemma 1 så finns det ett heltal x så att $x^2 \equiv -1 \pmod p$. Fixera ett sådant x och betrakta funktionen $f : \mathbf{Z} \to \mathbf{Z}$ som ges av

$$f(u, v) = u + xv.$$

Låt $K = [\sqrt{p}]$ så att $K < \sqrt{p} < K + 1$. Det finns $(K + 1)^2 > p$ par (u, v) av heltal så att $0 \leq u, v \leq K$. Alltså, enligt lådprincipen, måste det finnas två olika par $(u_1, v_1), (u_2, v_2)$ så att

$$f(u_1, v_1) \equiv f(u_2, v_2) \pmod p \Rightarrow (u_1 - u_2) \equiv -x(v_1 - v_2) \pmod p.$$

Låt $a := u_1 - u_2, b := v_1 - v_2$. Eftersom $x^2 \equiv -1 \pmod p$, har vi då att $a^2 + b^2 \equiv 0 \pmod p$. Minst ett av $a, b \neq 0$ - annars skulle $(u_1, v_1) = (u_2, v_2)$ - så att $a^2 + b^2 \neq 0$. Men eftersom alla u_i, v_i ligger i intervallen $[0, K]$ så måste då både a och b ligga i intervallen $[-K, K]$, så att $a^2 + b^2 \leq 2K^2 < 2p$.

Därför, har vi bevisat att $a^2 + b^2$ är en multipel av p, och ligger strängt mellan 0 och $2p$. Det följer att $a^2 + b^2 = p$.

References

[1] J. Silverman, The Arithmetic of Elliptic Curves. Springer 1986 (GTM No. 106).

Lecture 5-6 : 151104

Let n be a positive integer. In these lectures, we are interested in determining the structure of $(\mathbf{Z}/n\mathbf{Z})^\times$, the multiplicative group of invertible integers modulo n, as an abelian group. The result is contained in Corollary 10 and Theorem 11 below, the proof of the latter of which will be quite technical.

We begin by observing that the additive group of the ring $\mathbf{Z}/n\mathbf{Z}$ is a cyclic group of order n.

Theorem 9 (Chinese Remainder Theorem) *Let*

$$n = \prod_{i=1}^{k} p_i^{\alpha_i} \tag{24}$$

be the prime factorisation of n. Then we have an isomorphism of rings

$$\mathbf{Z}/n\mathbf{Z} \cong \prod_{i=1}^{k} \mathbf{Z}/p_i^{\alpha_i}\mathbf{Z}. \tag{25}$$

PROOF : It is easy to see that the map which takes $a \in \mathbf{Z}/n\mathbf{Z}$ to the k-tuple

$$[a \pmod{p_i^{\alpha_i}}]_{i=1}^{k}$$

is a ring-isomorphism.

Remark On the level of additive groups, the isomorphism (25) is a special case of the following basic result

Fundamental Theorem of finite abelian groups *Let G be a finite abelian group of order n. Let n have prime factorisation as in (24). Then, for each $i = 1, ..., k$, G has a unique subgroup G_i of order $p_i^{\alpha_i}$, and G is the internal direct product of the G_i. Moreover, each G_i can be expressed as the internal direct product of cyclic groups, the set of whose orders is uniquely determined.*

Corollary 10 *With n as in (24), we have an isomorphism of abelian groups*

$$(\mathbf{Z}/n\mathbf{Z})^\times \cong \prod_{i=1}^{k} (\mathbf{Z}/p_i^{\alpha_i}\mathbf{Z})^\times. \tag{26}$$

PROOF : This follows directly from the ring-isomorphism (25), since an element of a direct product of rings is invertible if and only if each of its' components is invertible.

Hence we have reduced our problem of determining the structure of the group $(\mathbf{Z}/n\mathbf{Z})^\times$ to the case where n is a prime power. The full answer is now given by

Theorem 11 *(i) Om p är ett udda primtal, då är $(\mathbf{Z}/p^\alpha\mathbf{Z})^\times$ cyklisk av ordning $\phi(p^\alpha) = p^{\alpha-1}(p-1)$ för alla α.*

(ii) $(\mathbf{Z}/2\mathbf{Z})^\times = < 1 > \cong C_1$.
$(\mathbf{Z}/2^\alpha\mathbf{Z})^\times = < -1 > \times < 5 > \cong C_2 \times C_{2^{\alpha-2}}$, *för* $\alpha > 1$.

DEFINITION : Let p be an odd prime and $\alpha \geq 1$. An integer a such that $a \pmod{p^\alpha}$ is a generator of the cyclic group $(\mathbf{Z}/p^\alpha\mathbf{Z})^\times$ is called a *primitive root modulo p^α*.

SKETCH PROOF OF THEOREM 11 : We divide the proof of the theorem into three parts :

Part I : proof of (i) for $\alpha = 1$,
Part II : proof of (i) for $\alpha > 1$,
Part III : proof of (ii).

There are quite a lot of gory details in Parts II and III of this procedure, which we leave to the reader to fill out completely.

Proof of Part I : The fact that the group $(\mathbf{Z}/p\mathbf{Z})^\times$ is cyclic was already noted during the proof of Lemma 1. It is a special case of the fact that the multiplicative group of a finite field is cyclic. So it suffices to prove the latter. The key observation for this is

Lemma 2 *Let G be a finite abelian group, written multiplicatively. If, for every $n > 0$, G contains at most n elements satisfying $g^n = 1$, then G is cyclic.*

PROOF OF LEMMA : Suppose that G is not cyclic. Since a direct product of cyclic groups of co-prime orders is also cyclic, it follows from the Fundamental Theorem that there must exist a prime p such that G contains a non-cyclic Sylow p-subgroup G_p. But then G_p must in turn contain a subgroup isomorphic to $C_p \times C_p$. In this subgroup we have p^2 elements satisfying $g^p = 1$, and hence G contains at least as many solutions to the same equation. Since $p^2 > p$, we see that G does not satisfy the hypothesis of the lemma.

Now let $\mathbf{F}$ be a finite field and let $n > 0$. Any solution to the equation $g^n = 1$ in $\mathbf{F}^\times$ may also be considered as a root of the polynomial $x^n - 1$ in $\mathbf{F}$. But this polynomial, being of degree n, has at most n roots in $\mathbf{F}$. Hence we see that the group $\mathbf{F}^\times$ satisfies the hypothesis of Lemma 2, thus is cyclic. This completes *Part I.*

Proof of Part II : The proof follows Baker, p.23-24. Fix an odd prime p. Let g be a primitive root modulo p (which exists by *Part I*). We shall show that for an appropriate choice of an integer x, the integer $g + px$ is a primitive root modulo p^α for every $\alpha > 1$. It is required to choose x such that

$$(g + px)^d \equiv 1 \pmod{p^\alpha} \quad \Rightarrow \quad p^{\alpha-1}(p-1) \mid d.$$

Note that the order of $g + px$ modulo p^α divides $p^{\alpha-1}(p-1)$ a priori, since the order of any element in a group divides the group order.

First, for any choice of x, the fact that g is a primitive root modulo p already implies that $p - 1$ must divide d, since $g + px \equiv g \pmod{p}$ and so

$$(g+px)^d \equiv 1 \pmod{p^\alpha} \Rightarrow (g+px)^d \equiv 1 \pmod{p} \Leftrightarrow g^d \equiv 1 \pmod{p} \Leftrightarrow p-1 \mid d.$$

Since $g^{p-1} \equiv 1 \pmod{p}$, we have $g^{p-1} = 1 + py$ for some integer y. The binomial theorem states that

$$(g + px)^{p-1} = \sum_{i=0}^{p-1} \binom{p-1}{i} g^{p-1-i}(px)^i.$$

Modulo p^2 only the terms $i = 0, 1$ contribute, and we have that $(g+px)^{p-1} = 1 + pz$ where

$$z \equiv y + (p-1)g^{p-2}x \pmod{p}. \tag{27}$$

Since the coefficient of x in (27) is not divisible by p, we can choose x such that z is not divisible by p. We now claim that this is sufficient for $g+px$ to be the required primitive root. It needs to be shown that, if $\mathrm{GCD}(z,p)=1$, then

$$(1+pz)^{p^m} \equiv 1 \pmod{p^\alpha} \quad \Rightarrow \quad \alpha - 1 \leq m. \tag{28}$$

Once again, this follows immediately from the binomial theorem, which in this case states that

$$(1+pz)^{p^m} = \sum_{i=0}^{p^m} \binom{p^m}{i} (pz)^i.$$

Since $(z,p)=1$, one sees immediately that p^{m+1} is the highest power of p dividing the $i=1$ term. With a little more care one checks that, since p is odd, p^{m+2} divides each term for $i>1$. Hence

$$(1+pz)^{p^m} \equiv 1 + p^{m+1} \pmod{p^{m+2}}, \quad \text{for any } m>0. \tag{29}$$

And (28) follows immediately from (29). This completes *Part II.*

Proof of Part III : Similar to the proof of *Part II*, in particular the use of the binomial theorem. We omit details, but just note that, in order to prove the theorem for $\alpha > 3$ (it may be proven for $\alpha \leq 3$ by inspection), one writes $5 = 1 + 2^2$ and uses the binomial theorem to prove that

$$(1+2^2)^{2^{\alpha-2}} \equiv 1 \pmod{2^\alpha}, \tag{30}$$

$$(1+2^2)^{2^{\alpha-3}} \equiv 1 + 2^{\alpha-1} \pmod{2^\alpha}. \tag{31}$$

Eq. (30) implies that the cyclic subgroup of $(\mathbf{Z}/2^\alpha\mathbf{Z})^\times$ generated by 5 has order $2^{\alpha-2}$, and (31) implies that -1 is not an element of this subgroup. Then elementary group theory implies that $(\mathbf{Z}/2^\alpha\mathbf{Z})^\times$ is the internal direct product of the subgroups generated by -1 and 5.

Lecture 7-8 : 17-191104

Tidigare bevisade vi följande två fakta :

(1) kongruensen $x^2 \equiv -1 \pmod{p}$ har en lösning omm antingen $p = 2$ eller $p \equiv 1 \pmod{4}$.

(2) $x^2 + y^2 = p$ har en lösning omm antingen $p = 2$ eller $p \equiv 1 \pmod{4}$.

Dessa fakta var redan välkända mot slutet av 1600-talet, när Fermat var det stora namnet inom talteori (dvs inom matematik, modulo Newton osv !!). Det dröjade över 100 år innan arbetet av tre stora matematiker - Euler, Legendre och framför allt Gauβ - ledde till en stor generalisering av dessa två resultat, och öppnade vägen till vad som kallas nuförtiden för 'algebraisk talteori'.

Ovanstående (1) och (2) är speciella fall av följande två problem resp.:

PROBLEM 1 : Lös den allmäna kvadratiska kongruensen

$$ax^2 + bx + c \equiv 0 \pmod{n}. \tag{32}$$

PROBLEM 2 : Lös (i heltalen) den allmäna kvadratiska ekvationen i två variabler

$$ax^2 + bxy + cy^2 + dx + ey + f = 0. \tag{33}$$

Problem 2 kallas för *representationsproblemet* för *binära kvadratiska former*. Båda problemen kan, på ett naturligt sätt, ytterliggare delas upp i tre 'subproblem', nämligen

DEL A : ange allmäna och lätt verifierbara kriteria för *existens* av en lösning.

DEL B : ge effektiva algoritmer för att *hitta* en lösning.

DEL C : ge allmäna (beräknabara) formler för *antalet* lösningar. Notera att i Problem 1 räcker det att hitta alla lösningar bland en mängd av representanter för $\mathbf{Z}/n\mathbf{Z}$. Därför är antalet lösningar alltid ändligt så länge som en 'lösning' betyder en lösning modulo n. Vi adopterar denna konvention framöver.

Under kommande lektioner skall vi studera dessa två problem mycket noggrant. Vi börjar med det kvadratiska kongruensproblemet (Problem 1) och framför allt existensfrågan (Del A). Men först, är det bra att satisfiera oss att vi förstår linjära kongruenser

Proposition 7 *Kongruensen* $ax \equiv b \ (mod\ n)$ *har en lösning omm* $SGD(a, n)$ *delar* b*. I så fall finns det en unik lösning modulo* n/d*, eller* d *st. lösningar modulo* n*, där* $d := SGD(a, n)$*.*

Bevis : Följer från Theorem 1.

I nästa två propositionerna reducerar vi den allmäna kvadratiska kongruensen till en mer hanterlig form.

Proposition 8 *Kongruensen (32) är ekvivalent med kongruensen*

$$y^2 \equiv d \ (mod\ 4an), \tag{34}$$

där $y = 2ax + b$ *och* $d = b^2 - 4ac$*.*

Bevis : Kvadratkomplettering.

Denna proposition säger att det räcker att betrakta kongruenser av formen $x^2 \equiv a \pmod{n}$. Nu kommer huvudsteget i reduktionen. Det är ganska tekniskt, men det är resultatet som är viktigt :

Proposition 9 *(i)* $x^2 \equiv a \ (mod\ n)$ *har en lösning omm* $x^2 \equiv a \ (mod\ p^\alpha)$ *har en lösning för varje* $p^\alpha \| n$*.*

(ii) Om $a = p^i a_1$ *för något* $i < \alpha$ *och* $(a_1, p) = 1$*, ett nödvändigt villkor för existens av en lösning till* $x^2 \equiv a \ (mod\ p^\alpha)$ *är att* $2 \mid i$*. I så fall ges lösningarna av* $x = p^{i/2} x_1$ *där* $x_1^2 \equiv a_1 \ (mod\ p^{\alpha - i})$*.*

(iii) Om $p > 2$ *och* $(a, p) = 1$ *då har* $x^2 \equiv a \ (mod\ p^\alpha)$ *en lösning omm* $x^2 \equiv a \ (mod\ p)$ *har en lösning. Antalet lösningar är antingen* 0 *eller* 2*.*

(iv) $x^2 \equiv 1 \ (mod\ 2)$ *har den enda lösningen* $x = \{\bar{1}\}$*.*

(v) $x^2 \equiv 1 \ (mod\ 4)$ *har lösningarna* $x = \bar{1}, \bar{3}$ *och* $x^2 \equiv 3 \ (mod\ 4)$ *har inga lösningar.*

(vi) Om $\alpha \geq 3$ och $2 \dagger a$, då har $x^2 \equiv a \ (mod\ 2^\alpha)$ en lösning omm $a \equiv 1 (mod\ 8)$, dvs omm $a \equiv 5^{2\lambda} \ (mod\ 2^\alpha)$ för något $\lambda \geq 0$. I så fall finns det exakt 4 lösningar, nämligen $\pm x_0, \pm 5^{2^{\alpha-3}} x_0$ där $x_0 \equiv 5^\lambda \ (mod\ 2^\alpha)$.

Bevis : Lämnades som en övning. Man använder Corollary 10 och Theorem 11, och det hela är elementär gruppteori.

Nu har vi reducerat Problem 1 (Del A,B och C !) till att lösa kongruensen

$$x^2 \equiv a \ (\mathrm{mod}\ p), \tag{35}$$

där p är ett udda primtal och $(a,p) = 1$. Notera att Del C av problemet nu är trivialt - det finns antingen 0 eller 2 lösningar.

Varning !! Vi fuskar lite grann här. För att genomföra reduktionen av (32) till (35) måste man faktorisera talet n. Jag känner inget bra sätt att komma runt detta.

Notation : Låt p vara ett udda primtal och $(a,p) = 1$. Vi sätter

$$\left(\frac{a}{p}\right) := \begin{cases} 1, & \text{om } (p,a) = 1 \text{ och } (35) \text{ har en lösning,} \\ -1, & \text{om } (p,a) = 1 \text{ och } (35) \text{ har ingen lösning,} \\ 0, & \text{om } p|a. \end{cases} \tag{36}$$

Symbolen $\left(\frac{a}{p}\right)$ kallas för en *Legendre symbol.*

Del A av Problem 1 letar efter ett effektivt kriterium för existensen av en lösning till (35). Ett svar ges av

Eulers kriterium $\left(\frac{a}{p}\right) \equiv a^{p-1/2} \ (mod\ p)$.

Bevis : $\left(\frac{a}{p}\right) = 1 \Leftrightarrow a$ är en kvadrat i $(\mathbf{Z}/p\mathbf{Z})^\times$.

Från den praktiska vinkeln ger Eulers kriterium ett effektiv sätt att lösa existensfrågan : det räcker att kunna beräkna $a^{p-1/2}$ (mod p) och det finns snabba algoritmer för att utföra beräkningen av a^b (mod c), för godtyckliga heltal a, b, c. Den enklaste använder en bas-2 utveckling av potensen b och kallas för *square-and-multiply* algoritmen.

Nu kommer vi såsmåningom till Gauβ' reciprocitets lag. Den leder också till ett mycket effektivt sätt att lösa existensfrågan för (35), men dess största interesse är teoretiskt. Först, är det klart att det är ett mycket vackrare och djupare resultat än Euler's kriterium, men historien slutar inte där. Idag vet vi att Gauβ' lag är ett specialt fall av en 'reciprocitets lag' av Artin (1927) som gäller för godyckliga abelska talkropps utvidningar L/K[1]. I fallet $K = \mathbf{Q}$ och $[L : \mathbf{Q}] = 2$ får man Gauβ' lag. Artin's reciprocitets lag är ett av de stora resultaten i den del av algebraisk talteori som kallas för *class field theory*, eller mer precist *abelian class field theory*[2].

Under 1900-talet har flera olika formuleringar av 'class field theory' utvecklats och vilken formulering du lär dig beror på vilken bok du läser. För en formulering i termer av så-kallade *adèles* och *idèles* (kanske den mest kända) se t.ex. [1]. För en alternativ formulering i termer av *grupp kohomologi* se t.ex. [2]. Den senare har också en kapitel om historien av 'class field theory'.

Okej, tillbaka till jorden. Först har vi

Proposition 10 *Låt p vara ett udda primtal och a, b godtyckliga heltal. Då gäller att*

$$\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) = \left(\frac{ab}{p}\right).$$

BEVIS : Följer direkt från Eulers kriterium.

Man kan nu tänka sig att Proposition 10 reducerar beräkningen av Legendre symbolen $\left(\frac{a}{p}\right)$ till tre fall : $a = -1, a = 2$ och $a = q$, ett udda primtal. Man skulle kunna klaga att vi fuskar här igen, eftersom vi måste först faktorisera talet a. Men det kommer att visa sig att vi kan komma runt detta.

I varje fall, Lemma 1 eller Eulers kriterium ger svaret direkt för $a = -1$. För $a = 2$ så kan vi använda

Gauss' lemma *Låt p vara ett udda primtal. För varje $n \in \mathbf{Z}$, låt $[n]$ beteckna det unika talet som satisfierar $[n] \equiv n \ (mod\ p)$ och $-\frac{1}{2}p < [n] < \frac{1}{2}p$.*

[1] En kropp som är en ändlig utvidning av $\mathbf{Q}$ kallas för en *talkropp*. En utvidning av talkroppar L/K kallas för *abelsk* om den är Galois och Galois gruppen av L/K är abelsk.

[2] There is also a class field theory for non-abelian number field extensions, but it is much more difficult and highly incomplete.

Låt nu $(a,p)=1$ och sätt $a_j=[aj]$ för varje $j \in \mathbf{Z}$. Då är

$$\left(\frac{a}{p}\right) = (-1)^l \tag{37}$$

där

$$l = \text{ antalet } j \text{ så att } 1 \le j \le \tfrac{p-1}{2} \text{ och } a_j < 0.$$

Bevis : (Som i Baker, s.28-29). We evaluate the product

$$P := \prod_{j=1}^{(p-1)/2} aj \ (\text{mod } p)$$

in two different ways. First, by its' very definition,

$$P = a^{(p-1)/2}\left(\frac{p-1}{2}\right)! \equiv \left(\frac{a}{p}\right)\left(\frac{p-1}{2}\right)!, \tag{38}$$

by Euler's criterion. On the other hand, by definition we also have that

$$P \equiv \prod_{j=1}^{(p-1)/2} a_j.$$

Now I claim that, if $j \neq k$, then $a_j \not\equiv \pm a_k$ (mod p). For if $a_j \equiv \pm a_k$ then $aj \equiv \pm ak \Rightarrow p|a(j \mp k) \Rightarrow p|j \mp k$. But both j and k lie in the interval $[1, \frac{p-1}{2}]$, so if $j \neq k$, then $|j \mp k| \le 2 \cdot (\frac{p-1}{2}) = p-1 < p$, which makes it impossible for this quantity to be divisible by p.

Thus we've established our claim. This implies that the quantities $|a_j|$ are just a permutation of the numbers $1, 2, ..., \frac{p-1}{2}$, as j runs from 1 to $\frac{p-1}{2}$. By definition, l of them are negative. Hence

$$P \equiv \prod_{j=1}^{(p-1)/2} a_j = (-1)^l \cdot \left(\frac{p-1}{2}\right)! \tag{39}$$

But (37) follows immediately from (38) and (39).

Om vi specialiserar till $a=2$ i Gauss' lemma, så får vi följande

Proposition 11 *Låt p vara ett udda primtal. Då gäller att*

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} 1, & om\ p \equiv \pm 1\ (mod\ 8), \\ -1, & om\ p \equiv \pm 3\ (mod\ 8). \end{cases}$$

BEVIS : (Som i Baker, s.29). Let $j \in [1, \frac{p-1}{2}]$. Then

$$[2j] > 0 \Leftrightarrow 2j \leq \frac{p-1}{2} \Leftrightarrow j \leq \lfloor \frac{p-1}{4} \rfloor.$$

Hence, for $a = 2$, the quantity l in Gauβ' lemma is just

$$l = \frac{p-1}{2} - \lfloor \frac{p-1}{4} \rfloor. \tag{40}$$

It is now a short but tedious computation to verify that, for any odd number p, the HL of (40) is congruent to $(p^2-1)/8$ modulo 2. This and Gauβ lemma yield the desired result.

Då har vi kvar att beräkna $\left(\frac{q}{p}\right)$ där både p och q är udda primtal. Det stora resultatet är

Theorem 12 (Gauss' reciprocitets lag) *Låt p, q vara udda primtal. Då är*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{1}{4}(p-1)(q-1)}. \tag{41}$$

BEVIS : M.h.a. Gauss' lemma, som i boken, s.29-30. A condensed version follows.

By Gauss' lemma, $\left(\frac{p}{q}\right) = (-1)^l$ where l is the number of integers $x \in [1, \frac{q-1}{2}]$ such that $[px]_q < 0$. The latter inequality holds if and only if there is an integer y such that

$$-\frac{q}{2} < px - qy < 0. \tag{42}$$

Hence, l equals the number of integer solutions to the pair of inequalities (42) which in turn satisfy

$$0 < x < \frac{q}{2}. \tag{43}$$

Note that the right-hand inequality in (42) implies that $y > (\frac{p}{q})x > 0$, whereas the left-hand inequality, together with (43) imply that

$$qy < px + \frac{q}{2} < p\left(\frac{q}{2}\right) + \frac{q}{2} = \left(\frac{p+1}{2}\right) q \Rightarrow y < \frac{p+1}{2} \Rightarrow y < \frac{p}{2},$$

since y is an integer. In other words, every integer solution to (42) and (43) also satisfies

$$0 < y < \frac{p}{2}. \tag{44}$$

A similar analysis gives that $(\frac{q}{p}) = (-1)^m$, where m is the number of intgeer solutions to (43), (44) and the double-inequality (got by simultaneously interchanging $p \leftrightarrow q$, $x \leftrightarrow y$ in (42),(43) and (44))

$$0 < px - qy < \frac{p}{2}. \tag{45}$$

Hence the VL of (41) equals $(-1)^{l+m}$, where $l + m$ is the total number of integer solutions to (43), (44) and (got by combining (42) and (45))

$$-\frac{q}{2} < px - qy < \frac{p}{2}. \tag{46}$$

Eqs. (43) and (44) define a rectangle $\mathcal{R}$ containing $\frac{1}{4}(p-1)(q-1)$ integer points. Hence, to complete the proof, it suffices to show that there are an even number of integer points in this rectangle which do not satisfy (46). These points are contained in two disjoint subsets A and B of $\mathcal{R}$, where

$$A := \{(x, y) \in \mathcal{R} : px - qy < -p/2\},$$
$$B := \{(x, y) \in \mathcal{R} : px - qy > q/2\}.$$

To prove that the number of integer points in $A \cup B$ is even, it suffices to establish a 1-1 correspondence between the integer points in A and those in B. One may now tediously verify that such a correspondence is given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \leftrightarrow \begin{pmatrix} \frac{1}{2}(q+1) - x \\ \frac{1}{2}(p+1) - y \end{pmatrix}.$$

Anmärkning : Bokens bevis är ganska elementärt och mycket snyggt, och är ganska nära Gauβ' ursprunlgiga idéer eftersom det använder hans lemma. Under sitt liv upptäckte Gauβ minst 8 olika bevis av sitt lag[3]. Några av dessa har en lite mer analytisk inriktning, och beror på beräkningen av en typ av ändlig summa som kallas för en *Gauβ summa*. Summor av denna typ har visat sig uppstå i många olika sammanhang inom talteori (både algebraisk och analytisk), t.ex. i beviset av Dirichlet's sats som kommer senare under kursen. Vi skall diskutera dem nästa gång och presentera ytterligare ett bevis av reciprocitetslagen.

References

[1] S. Lang, ALgebraic number theory, Springer.

[2] J. Cassels and A. Fröhlich eds., Algebraic number theory, Academic Press 1967.

[3]Moderna böcker betraktas innehålla ungefär 40 olika bevis. Men Urban Larsson påstår att det finns 196 stycken på nätet. Isn't technology wonderful !

Lectures 9-10 : 22-241104

In these lectures we introduce the notions of *group character* and *Gauss sum*. An immediate application will be an alternative proof of Gauss' reciprocity law. Later in the course, we shall find further application of these ideas during the proof of Dirichlet's Theorem.

NOTATION : $\mathbf{C}^\times$ denotes the group of non-zero complex numbers under multiplication. All groups in this section are written multiplicatively.

DEFINITION : Let G be a group. A *character* of G is a group homomorphism

$$\chi : G \to \mathbf{C}^\times.$$

The First Isomorphism Theorem in group theory implies that, for any character χ,

$$\frac{G}{\mathrm{Ker}\chi} \cong \mathrm{Im}\chi.$$

Since Im $\chi \subseteq \mathbf{C}^\times$ is abelian, we find that Ker $\chi \supseteq G'$, the commutator subgroup of G. Hence the characters of G are identical to those of the abelian group G/G'. It thus suffices to study characters of abelian groups. In this course, we shall only study characters of finite abelian groups.

Note that, when G is finite, $\chi(g)$ must be a root of unity for any $g \in G$ and any character χ. Indeed, if $g^n = 1_G$, then $\chi(1_G) = 1 = \chi(g^n) = [\chi(g)]^n$, so $\chi(g)$ is an n:th root of unity.

Given two characters χ_1, χ_2 of a group G, we can define a third character $\chi_1 \cdot \chi_2$, called the 'product' of χ_1 and χ_2, as follows :

$$(\chi_1 \cdot \chi_2)(g) := \chi_1(g) \cdot \chi_2(g), \qquad \forall\, g \in G. \tag{47}$$

It is easy to check that $\chi_1 \cdot \chi_2$ is also a character.

NOTATION : The set of characters of an abelian group G is denoted by $\hat{G}$.

Theorem 13 *Let G be a finite abelian group. Under the multiplication of characters defined by (47), $\hat{G}$ becomes a finite abelian group isomorphic to G.*

PROOF : The details were given in class. To show that $\hat{G}$ is a group (that the multiplication is associative, and that an identity element and inverses exist) is easy. That this group is isomorphic to G is the interesting part. The idea for showing this was as follows : Let

$$G = C_1 \oplus C_2 \oplus \cdots \oplus C_k$$

be any decomposition of G as a direct sum of cyclic groups. Let n_i denote the order of C_i and let e_i be any group element which generates C_i. We think of the set $\{e_1, ..., e_k\}$ as a 'basis' for G. Then there is a canonical 'dual basis' $\{\chi_1, ..., \chi_k\}$ for $\hat{G}$ given by

$$\chi_i(e_j) = \begin{cases} e^{2\pi i/n_i}, & \text{if } i = j, \\ 1, & \text{if } i \neq j. \end{cases}$$

NOTATION/TERMINOLOGY : The identity element of the character group $\hat{G}$ is denoted χ_0. It is called the *trivial character* and is the mapping given by

$$\chi_0(g) = 1, \quad \forall\, g \in G.$$

The following proposition will prove useful in our study of Dirichlet L-functions.

Proposition 12 *Let G be a finite abelian group, $g^* \in G$ and $\chi^* \in \hat{G}$. Then*

(i)

$$\sum_{\chi \in \hat{G}} \chi(g^*) = \begin{cases} |G|, & \text{if } g^* = 0, \\ 0, & \text{if } g^* \neq 0. \end{cases}$$

(ii)

$$\sum_{g \in G} \chi^*(g) = \begin{cases} |G|, & \text{if } \chi^* = \chi_0, \\ 0, & \text{if } \chi^* \neq \chi_0. \end{cases}$$

PROOF : We present the proof for (i) ; that for (ii) is similar. The argument is as follows : if $g \neq 1$ then there exists at least one character - let's pick one and denote it χ_g - with the property that $\chi_g(g) \neq 1$. This follows from the proof of Theorem 13, as outlined above.

Then, since $\hat{G}$ is a group, we have that

$$\sum_{\chi} \chi(g) = \sum_{\chi} (\chi \cdot \chi_g)(g) = \chi_g(g) \cdot \sum_{\chi} \chi(g).$$

Since $\chi_g(g) \neq 1$, it follows that the sum must be zero, v.s.v.

Definition : Let n be a positive integer. A *character modulo n* is a mapping

$$\chi : \mathbf{Z}/n\mathbf{Z} \to \mathbf{C}$$

such that

(i) $\chi(a) = 0$ if $\mathrm{SGD}(a, n) > 1$,
(ii) the restriction of χ to $(\mathbf{Z}/n\mathbf{Z})^{\times}$ is a multiplicative group character.

Example : If p is an odd prime, then

$$\chi(a) := \left(\frac{a}{p}\right),$$

the Legendre symbol, is a character modulo p. It is the only non-trivial *real* character modulo p, i.e.: the only non-trivial character modulo n such that $\mathrm{Im}\ \chi \subseteq \mathbf{R}$.

Notation : Let r, s be arbitrary real numbers, $s \neq 0$. We denote

$$e^{2\pi i \frac{r}{s}} := e_s(r).$$

Definition : Let n be a positive integer and χ a character modulo n. The *Gauss sum* for χ is defined as

$$G(\chi) := \sum_{m=0}^{n-1} \chi(m) e_n(m).$$

Theorem 14 *Let p be an odd prime, χ be the character modulo p induced by the Legendre symbol (as above), and set $G := G(\chi)$. Then*

$$G^2 = \left(\frac{-1}{p}\right) p = \begin{cases} p, & \text{if } p \equiv 1 \ (\text{mod } 4), \\ -p, & \text{if } p \equiv 3 \ (\text{mod } 4). \end{cases}$$

PROOF : Explicitly we have

$$G^2 = \sum_{m_1=0}^{p-1} \sum_{m_2=0}^{p-1} \left(\frac{m_1}{p}\right)\left(\frac{m_2}{p}\right) e_p(m_1) e_p(m_2)$$
$$= \sum_{m_1=1}^{p-1} \sum_{m_2=1}^{p-1} \left(\frac{m_1 m_2}{p}\right) e_p(m_1+m_2).$$

In the inner sum we make a change of variables

$$m_2 \equiv m_1 n \ (\mathrm{mod}\ p).$$

Observe that, since p is prime, as m_2 runs over all non-zero residue classes mod p (for a fixed m_1), the same is true of n. Hence we obtain that

$$G^2 = \sum_{m_1=1}^{p-1} \sum_{n=1}^{p-1} \left(\frac{m_1 \cdot m_1 n}{p}\right) e_p(m_1 + m_1 n)$$
$$= \sum_{m_1=1}^{p-1} \sum_{n=1}^{p-1} \left(\frac{n}{p}\right) e_p[m_1(1+n)].$$

We now reverse the order of summation, thus finding that

$$G^2 = \sum_{n=1}^{p-1} \left(\frac{n}{p}\right) \sum_{m_1=1}^{p-1} e_p[m_1(1+n)]. \tag{48}$$

At this point we insert

Lemma 3 *Let* $\mathbf{F}$ *be any field,* p *be a prime and* ξ *a p:th root of unity (i.e.:* $\xi^p = 1$*) in any extension field of* $\mathbf{F}$*. Then*

$$\sum_{i=0}^{p-1} \xi^i = \begin{cases} p, & \textit{if } \xi = 1, \\ 0, & \textit{if } \xi \neq 1. \end{cases}$$

PROOF OF LEMMA : It is obvious that the sum is p if $\xi = 1$. Otherwise ξ^i runs over all the roots of the polynomial $x^p - 1$, and the sum is just the sum of the roots. But this sum must then equal the coefficient of x^{p-1}, which is zero.

Now back to the proof of Theorem 14. Lemma 3 implies that

$$\sum_{m_1=1}^{p-1} e_p[m_1(1+n)] = \begin{cases} p-1, & \text{if } n = p-1, \\ -1, & \text{otherwise.} \end{cases}$$

Substituting into (48), we obtain that

$$\begin{aligned} G^2 &= -\sum_{n=1}^{p-2} \left(\frac{n}{p}\right) + (p-1)\left(\frac{p-1}{p}\right) \\ &= p \cdot \left(\frac{-1}{p}\right) - \sum_{n=1}^{p-1} \left(\frac{n}{p}\right). \end{aligned}$$

But Proposition 12(ii) implies that the last sum is zero, and this completes the proof of Theorem 14.

Remark 1 From Theorem 14 we deduce immediately that $|G| = \sqrt{p}$ and, more precisely, that

$$G = \begin{cases} \pm\sqrt{p}, & \text{if } p \equiv 1 \pmod 4, \\ \pm i\sqrt{p}, & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

In fact, the + sign is correct in both cases, but to prove this seems to be quite difficult. It is usually done with the help of some Fourier analysis.

Remark 2 Let $n > 0$ be any integer, not necessarily prime. A character χ modulo n is said to be *primitive* if there is no proper divisor d of n such that the function χ is periodic with period d. It turns out that

$$|G(\chi)| = \sqrt{n},$$

for any $n > 0$ and any primitive character χ modulo n. For a proof, see Chapter 9 of [1].

We now apply Theorem 14 to give an alternative proof of the Guass reciprocity law. So let p and q be distinct odd primes and suppose WLOG that $p < q$. We shall work in the field $\mathbf{F} := \mathbf{F}_{p^{q-1}}$. By Fermat's Theorem, $p^{q-1} \equiv 1 \pmod q$, so the order of the multiplicative group $\mathbf{F}^{\times}$ is divisible by q. Hence $\mathbf{F}$ contains a primitive q:th root of unity, which we denote by ξ. We shall consider the (generalised) Gauss sum

$$G := \sum_{m=0}^{q-1} \left(\frac{m}{q}\right) \xi^m.$$

We shall evaluate G^p in two different ways, which will immediately yield the reciprocity law.

First evaluation :

We write

$$G^p = \left(G^2\right)^{\frac{p-1}{2}} G.$$

Now because Lemma 3 still holds in this setting, the proof of Theorem 14 goes through verbatim and we find that

$$G^2 = \left(\frac{-1}{q}\right) q = (-1)^{\frac{q-1}{2}} q.$$

Hence

$$G^p = \left[(-1)^{\frac{q-1}{2}} q\right]^{\frac{p-1}{2}} G = (-1)^{\frac{1}{4}(p-1)(q-1)} q^{\frac{p-1}{2}} G. \tag{49}$$

Second evaluation :

In a field of characteristic p, the binomial theorem immediately implies that, for any two field elements a and b,

$$(a+b)^p = a^p + b^p.$$

Now our field $\mathbf{F}$ has characteristic p, so applying the above to our Gauss sum, we find that

$$\begin{aligned} G^p &= \sum_{m=1}^{q-1} \left(\frac{m}{q}\right)^p \xi^{mp} \\ &= \sum_{m=1}^{q-1} \left(\frac{m}{q}\right) \xi^{mp}. \end{aligned}$$

Now we change variables from m to mp. Note that as m runs over all non-zero residue classes mod q, the same is true of mp. We abuse notation and write p^{-1} instead of $p^{-1} \pmod q$. Thus we have

$$G^p = \sum_{m=1}^{q-1} \left(\frac{mp^{-1}}{q}\right) \xi^m \tag{50}$$

$$= \left(\frac{p^{-1}}{q}\right) \sum_{m=1}^{q-1} \left(\frac{m}{q}\right) \xi^m$$

$$= \left(\frac{p}{q}\right) G.$$

Equating (49) and (50) we obtain that, as elements of the finite field $\mathbf{F}$,

$$(-1)^{\frac{1}{4}(p-1)(q-1)} q^{\frac{p-1}{2}} G = \left(\frac{p}{q}\right) G.$$

We can cancel G from this equation since Theorem 14 implies that $G \neq 0$. Also, Euler's criterion states that

$$\left(\frac{q}{p}\right) \equiv q^{\frac{p-1}{2}} \pmod{p}.$$

In the field $\mathbf{F}$, which has characteristic p, this congruence becomes an equality. We have thus shown that, in $\mathbf{F}$,

$$(-1)^{\frac{1}{4}(p-1)(q-1)} \left(\frac{q}{p}\right) = \left(\frac{p}{q}\right).$$

But we can also consider this as an equation between ordinary integers, in which case it immediately gives the reciprocity law. The proof is complete.

We close this lecture by defining an extension of the Legendre symbol, whose properties will enable us to describe a fast algorithm for computing Legendre symbols (as an alternative to using Euler's criterion).

Definition : Let n be any odd integer, and let

$$n = \prod_{i=1}^{k} p_i^{\alpha_i}$$

be its' prime factorisation. For any integer a, we define the *Jacobi symbol* $\left(\frac{a}{n}\right)$ by

$$\left(\frac{a}{n}\right) := \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)^{\alpha_i},$$

where each of the terms on the HL is an ordinary Legendre symbol.

Obs! Note that $\left(\frac{a}{n}\right) = 0$ if and only if $\mathrm{SGD}(a, n) > 1$. Otherwise, $\left(\frac{a}{n}\right) = \pm 1$. The Jacobi symbol is multiplicative, as in Proposition 10 ; in particular, it defines a real character modulo n. However, if n is not a prime, it is not necessarily the case that $\left(\frac{a}{n}\right) = 1 \Leftrightarrow a$ is a quadratic residue modulo n. Indeed, by Proposition 9, a is a quadratic residue mod n if and only if $(\frac{a}{p}) = 1$ for every prime p dividing n. But if, for example, n is a product of two distinct primes p and q, then $(\frac{a}{p}) = (\frac{a}{q}) = -1 \Rightarrow (\frac{a}{n}) = +1$.

We have the following extension of Euler's criterion, Proposition 11 and Theorem 12 :

Theorem 15 *Let m, n be any two odd integers. Then*
(i)

$$\left(\frac{-1}{n}\right) = (-1)^{\frac{n-1}{2}} = \begin{cases} 1, & \textit{if } n \equiv 1 \ (\textit{mod } 4), \\ -1, & \textit{if } n \equiv 3 \ (\textit{mod } 4). \end{cases}$$

(ii)

$$\left(\frac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}} = \begin{cases} 1, & \textit{if } n \equiv \pm 1 \ (\textit{mod } 8), \\ -1, & \textit{if } n \equiv \pm 3 \ (\textit{mod } 8). \end{cases}$$

(iii)

$$\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) = (-1)^{\frac{1}{4}(m-1)(n-1)}.$$

PROOF : The proofs of the various parts of Theorem 15 employ the corresponding results for Legendre symbols, the definition of the Jacobi symbol and repeated use of the fact that, if n_1, n_2 are any two odd integers, then

$$\frac{1}{2}(n_1 - 1) + \frac{1}{2}(n_2 - 1) \equiv \frac{1}{2}(n_1 n_2 - 1) \ (\text{mod } 2).$$

We omit the mind-numbingly boring details.

Referemces

[1] H. Davenport, Multiplicative Number Theory. Springer.

Lecture 11-12 : 2911-011204

Idag börjar vi studiet av kvadratiska ekvationer över $\mathbf{Z}$, dvs ekvationer av formen

$$f(x_1, ..., x_n) = 0, \tag{51}$$

där f är ett polynom av grad 2 med heltals koefficienter. Vi söker heltalslösningar till (51).

För $n = 1$ är allting klart. Ekvationen

$$ax^2 + bx + c = 0 \qquad (a \neq 0) \tag{52}$$

har lösningar i $\mathbf{C}$ som ges av

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{53}$$

Alltså, har ekvationen en lösning i $\mathbf{Q}$ omm $b^2 - 4ac = d_1^2$ är en perfekt kvadrat, och en lösning i $\mathbf{Z}$ omm $2a \mid -b \pm d_1$. Antalet lösningar i $\mathbf{Z}$ är då $0, 1$ eller 2.

För $n = 2$ blir situationen redan mycket mer komplicerad att analysera. I den här kursen ska vi koncentrera mest på detta fall.

Det var framför allt Gauβ som utvecklade en allmän teori för kvadratiska ekvationer i två variabler. Den större delen av hans klassiska arbete *Disquisitiones Arithmeticae* handlar om denna teori. I vår presentation ska vi, till en början, följa Gauβ. Om tiden räcker så kommer vi att reformulera våra resultat i mer modern algebraisk talteoretiskt språk.

Vi studerar då den allmäna kvadratiska ekvationen i två variabler, dvs en ekvation av formen

$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \tag{54}$$

där koefficienterna $a, b, c, d, e, f \in \mathbf{Z}$ och $(a, b, c) \neq (0, 0, 0)$. Polynom som på vänster sidan av (54) ska betecknas framöver med $f(x, y), g(x, y)$ mm. Det är ofta passande att använda en matris notation. Om vi låter

$$A := \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}, \quad g := \begin{pmatrix} d \\ e \end{pmatrix}, \quad X := \begin{pmatrix} x \\ y \end{pmatrix}, \tag{55}$$

då kan (54) skrivas på formen

$$f(x,y) = f(X) = X^T AX + g^T X + f = 0. \tag{56}$$

För att simplificera formen på ekvationen lite, betraktar man koordinat bytningar av typen $X \to X'$ där

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = X' := UX + V, \qquad \exists\, U \in GL_2(\mathbf{Q}),\ V \in \mathbf{Q}^2. \tag{57}$$

Om $\det(A) \neq 0$ då kontrollerar man lätt att transformationen $U = I, v = -\frac{1}{2}(A^T)^{-1}g$ skaffar en ekvation på formen

$$f'(x',y') = a(x')^2 + bx'y' + c(y')^2 = n. \tag{58}$$

Här är kanske n inte ett heltal, men då behöver man bara klara nämnarna.

Fallet $\det(A) = 0$ är speciellt. Här ser man, efter en kvadrat komplettering (se (64) nedan), att det finns en koordinat bytning som tar (54) till formen

$$(x')^2 + e'y' + f' = 0. \tag{59}$$

Denna ekvation är tämligen ointeressant eftersom det är klart att det finns två heltalslösningar för varje y' så att $-e'y' - f'$ är en perfekt kvadrat.

Alltså, i fortsättningen ska vi ägna oss åt att studera ekvationer av formen (58), där $b^2 - 4ac \neq 0$.

Definition 1 : En funktion $f : \mathbf{Z}^2 \to \mathbf{Z}$ som ges av

$$f(x,y) = ax^2 + bxy + cy^2, \qquad a,b,c \in \mathbf{Z}, \tag{60}$$

kallas för en *(integral) binär kvadratisk form.* I matris form, ges f av

$$f(X) = X^T AX, \quad X = \begin{pmatrix} x \\ y \end{pmatrix}, \ A = \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}. \tag{61}$$

Definition 2 : Kvantiteten $d := b^2 - 4ac$ kallas för *diskriminanten* av formen f. Notera att

$$d_f = -4\det(A_f). \tag{62}$$

Proposition 13 *An integer d is the discriminant of some binary quadratic form iff $d \equiv 0$ or 1 (mod 4).*

PROOF : The definition of d immediately implies the necessity. For sufficiency, it suffices to write down explicit forms, and we choose

$$\begin{array}{ll} x^2 - \frac{1}{4}dy^2, & \text{if } d \equiv 0 \pmod 4, \\ x^2 + xy - \frac{1}{4}(d-1)y^2, & \text{if } d \equiv 1 \pmod 4. \end{array} \tag{63}$$

The forms in (63) are called the *principal forms.*

DEFINITION 3 : En form kallas för *degenererad* om $d = 0$; annars kallas den för *icke-degenererad.* En form med $d < 0$ kallas för *definit* : *positiv definit* i fallet $a > 0$ och *negativ definit* i fallet $a < 0$. En form med $d > 0$ kallas för *indefinit.* Slutligen, kallas formen för *faktoriserbar* om d är en perfekt kvadrat.

Notera att f är positiv definit omm $-f$ är negativ definit. Då får vi strunta i negativa former framöver.

DEFINITION 4 : Låt $n \in \mathbf{Z}$. En form f sägs *representera* n om $f(x, y) = n$ för något $(x, y) \in \mathbf{Z}^2$. En sådan representation av n kallas för *egentlig (proper)* om $\gcd(x, y) = 1$.

Problemet att lösa (58) i heltalen kallas för *representationsproblemet* för binära kvadratiska former. Det är det grundläggande problemet som motiverar alla idéerna som vi skall nu presentera.

Först förklarar vi terminologin i Definition 3. Kvadratkomplettering i (60) leder till uttrycket

$$4af(x, y) = (2ax + by)^2 - dy^2. \tag{64}$$

Då ser vi direkt att representationsproblemet är tämligen ointeressant för faktoriserbara former, eftersom det reduceras till att faktorisera $4an$ och att då lösa par av simultana LINJÄRA ekvationer. I fortsättningen, alltså, antar vi alltid följande

‘*f är en kvadratisk form vars diskriminant d inte är en perfekt kvadrat, och om $d < 0$ då är f positiv definit*’.

Nu leder (64) direkt till följande resultat som förklarar vår
terminologi :

Proposition 14 *(i) f positiv definit $\Leftrightarrow$ f representerar bara positiva tal. (ii) f indefinit $\Leftrightarrow$ f representerar både positiva och negativa tal.*

Gauβ' approach to the representation problem can be divided up into the following main steps :

STEP 1 : Divide up all the forms according to their discriminant.

STEP 2 : Define an equivalence relation on the set of forms of a given discriminant so that equivalent forms need not be distinguished as regards the representation problem, i.e.: equivalent forms represent (properly) the same integers and there is a canonical 1-1 correspondence between their (proper) representations of a given integer.

STEP 3 : Show that the number of equivalence classes for a given discriminant is finite and that each class contains at least one 'nice' form.

STEP 4 : Find a formula for the total number of representations of an integer by a representative collection of forms of a given discriminant.

We shall carry out this program. But first, some remarks are in order :

1. The process described in Steps 2-3 is called *reduction theory.* It turns out that a satisfactory reduction theory is far easier to obtain for definite forms than for indefinite forms. We will only present detailed proofs of the results in the former case. In both cases, the theory gives an algorithm for deciding whether two forms of a given discriminant are equivalent. The procedure is simpler in the definite case.

2. With regard to Step 4, the formula in question will be seen to be quite simple and elegant. In the case of indefinite froms, where any given integer may have infinitely many representations as we'll see below, the formula counts representations of a certain type, called *primary.* It is important to note that there seems to be no such simple formula for the number of (primary) representations of an integer by a single form of a given discriminant. However, we will see that there exist, in principle, algorithms for deciding

whether (58) has a solution, and for counting the number of (primary) solutions. For definite forms, this is in fact a trivial result, since $f(x, y) = n$ implies that x and y are bounded explicitly. But for indefinite forms, this is not the case and something more clever must be done. Also, the algorithms for definite forms got by explicitly bounding x, y are obviously very slow. Finally, we observe that our algorithms here will in turn depend on those for deciding equivalence of forms, as just mentioned above.

DEFINITION : The binary quadratic forms f and f' are said to be *equivalent*, denoted $f \sim f'$, if there exists a matrix $M \in SL_2(\mathbf{Z})$ such that

$$M^T A_f M = A_{f'}. \tag{65}$$

It is readily checked that $\mathrm{SL}_2(\mathbf{Z})$ is a group, and hence that (65) does indeed define an equivalence relation on the set of all binary forms, and that equivalent forms have the same discriminant. The importance of the relation lies in the following fact :

Proposition 15 *If $f \sim f'$, then for each integer n, there is a 1-1 correspondence between the (proper) representations of n by f and f'.*

BEVIS : If $M^T A_f M = A_{f'}$ then

$$f'(X) = X^T A_{f'} X = n \Leftrightarrow f(MX) = (X^T M^T) A_f (MX) = n.$$

Invertibility of M inside $\mathrm{SL}_2(\mathbf{Z})$ implies that the correspondence $X \leftrightarrow MX$ between representations of n is 1-1, and that proper representations correspond to proper representations.

It is convenient to write down, once and for all, the explicit relationship between the coefficients of equivalent forms. If $f = \{a, b, c\}$, $f' = \{a', b', c'\}$ and $M = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ is a matrix taking f to f', then (65) gives

$$\begin{aligned} a' &= f(p, r) = ap^2 + bpr + cr^2, \\ c' &= f(q, s) = aq^2 + bqs + cs^2, \\ b' &= 2apq + b(ps + qr) + 2crs. \end{aligned} \tag{66}$$

Remark It is not hard to show that the centre of the group $\mathrm{SL}_2(\mathbf{Z})$ consists of just $\{\pm I\}$. The quotient group $\mathrm{SL}_2(\mathbf{Z})/\{\pm I\}$ is denoted $\mathrm{PSL}_2(\mathbf{Z})$ (P for *projective*). It is a little harder to prove that $\mathrm{PSL}_2(\mathbf{Z})$ is a *simple* group, i.e.: it possesses no proper normal subgroups. Note that, in (65), M and $-M$ define the same transformation of any form f. Hence the action of the group $\mathrm{SL}_2(\mathbf{Z})$ on the set of all binary quadratic forms defined by (65) is, in reality, an action of $\mathrm{PSL}_2(\mathbf{Z})$. The most important group-theoretic property of $\mathrm{SL}_2(\mathbf{Z})$ for our subsequent investigations is

Theorem 16 *$SL_2(\mathbf{Z})$ is generated by the two elements*

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

PROOF : Let X denote the subgroup of $\mathrm{SL}_2(\mathbf{Z})$ generated by S and T. Suppose X is not the whole of $\mathrm{SL}_2(\mathbf{Z})$, and let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a matrix in the complement of X. We must derive a contradiction from the supposed existence of M.

The first step is to show that both entries in the left-hand column of M must be different from zero. For suppose $c = 0$. Then $\det(M) = ad = 1 \Rightarrow a = d = \pm 1$. Thus M has the form $\pm \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$, for some integer b, i.e.: $M = \pm T^b$. But one may check that $S^2 = -I$ so $M = T^b$ or $M = S^2 T^b$. In both cases, $M \in X$, a contradiction.

Similarly, suppose $a = 0$. Then $SM = \begin{pmatrix} -c & -d \\ 0 & b \end{pmatrix}$ has a zero in the $(2,1)$-position, hence lies in X, by the above. But then $M \in X$ also. This completes the first step of the proof.

Choose now the matrix M such that $\min\{|a|, |c|\}$ is minimal. We will derive a contradiction by producing another matrix M' outside X such that $\min\{|a'|, |c'|\} < \min\{|a|, |c|\}$.

WLOG, we may assume that $|a| \geq |c|$, since left-multiplication by S interchanges $|a|$ and $|c|$. Let now q, r be the uniquely defined integers satisfying $a = cq + r$ and $0 \leq r < c$. Then set

$$M' := T^{-q} M = \begin{pmatrix} 1 & -q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} r & b - qd \\ c & d \end{pmatrix}.$$

Clearly, M' lies outside X if M does, and

$$\min\{|a'|, |c'|\} = \min\{r, |c|\} = r < |c| = \min\{|a|, |c|\}, \quad \text{v.s.v..}$$

Remark $S^4 = I$ whereas T is clearly an element of infinite order in $\mathrm{SL}_2(\mathbf{Z})$.

Lektion 13-14 : 02-031204

DEFINITION : The form $\{a, b, c\}$ is said to be *reduced* if either

$$-|a| < b \le |a| < |c| \quad \text{or} \quad 0 \le b \le |a| = |c|. \tag{67}$$

The main result of reduction theory is

Theorem 17 (Lagrange/Gauss). *(i) Every binary form is equivalent to a reduced form.*
(ii) There are only finitely many reduced forms of a given discriminant.
(iii) Every positive definite form is equivalent to precisely one reduced form.

PROOF : A condensed version of the proof to follow is found on ps. 36-7 of Baker's book.

PROOF OF PART (I) : Let $f = \{a, b, c\}$ be any form. Thus $A_f = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$. By Theorem 16, we need to show that there is a sequence of transformations by S and T which takes A_f to a form satisfying (67). One readily computes the effect of transformations by S and T as (see (66)) :

$$S^T \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} S = \begin{pmatrix} c & -b/2 \\ -b/2 & a \end{pmatrix}, \tag{68}$$

$$(T^k)^T \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} T^k = \begin{pmatrix} a & (b+2ak)/2 \\ (b+2ak)/2 & ak^2+bk+c \end{pmatrix}, \quad \text{for any } k \in \mathbf{Z}. \tag{69}$$

Eq. (68) says that one should transform by S if $|a| > |c|$. Eq. (69) says that one should transform by a suitable power of T if $b \notin (-|a|, |a|]$. Since each S-transformation reduces the absolute value of the $(1,1)$-entry in A_f, it is thus clear that there is a finite sequence of transformations which takes A_f to a form satisfying $|a| \le |c|$ and $b \in (-|a|, |a|]$. Finally, if now $|a| = |c|$ and $b < 0$, then a final transformation by S retains $|a| = |c|$ this while making $b \ge 0$.

PROOF OF PART (II) : From (67) and the equation $d = b^2 - 4ac$, we can easily derive explicit bounds for the a-coefficient of a reduced form of discriminant d, namely

$$0 < |a| \le \sqrt{-d/3}, \quad \text{for a reduced definite form,} \tag{70}$$

$$0 < |a| \leq \sqrt{d/4}, \quad \text{for a reduced indefinite form.} \tag{71}$$

The boring details of these computations are left to the reader. Now since $|a|$ is bounded, thee are only finitely many possibilities for a. But a reduced form satisfies $|b| \leq |a|$, so there are also only finitely many possibilities for b. But a, b and d uniquely determine c, so there are in turn only finitely many possible values of c, and hence in all only finitely many possibilities for the reduced form $\{a, b, c\}$.

PROOF OF PART (III) : Let $f = \{a, b, c\}$ and $f' = \{a', b', c'\}$ be two reduced positive definite forms of the same discriminant $d < 0$. Suppose that $f \sim f'$. We must show that $a = a'$, $b = b'$ and $c = c'$. That the forms are positive definite and reduced implies that

$$-a < b \leq a < c \quad \text{or} \quad 0 \leq b \leq a = c,$$

and similarly for the primed values.

The first step is to show that if $f = \{a, b, c\}$ is reduced and positive definite then

(i) a is the smallest integer represented properly by f
(ii) either $a = c$ or c is the second smallest integer properly represented by f.

Note that (i) and (ii) immediately imply that $a = a'$ and $c = c'$, since equivalent forms represent the same numbers. We now establish these two facts.

Direct computations show that $f(\pm 1, 0) = a$ and $f(0, \pm 1) = c$. Hence both a and c are properly represented by f. Let now (x, y) be any other integer pair with $\mathrm{GCD}(x, y) = 1$. Then both x and y have to be non-zero. Suppose WLOG that $|x| \geq |y|$. We have

$$\begin{aligned} f(x, y) &= ax^2 + bxy + cy^2 \\ &\geq a|x|^2 - |b||x||y| + c|y|^2 \\ &\geq a|x|^2 - |b||x|^2 + c|y|^2 \\ &= (a - |b|)|x|^2 + c|y|^2 \\ &\geq a - |b| + c \geq c, \end{aligned}$$

where the last two inequalities follow from the facts that $a \geq |b|$, the form being reduced, and that $|x| \geq 1$, $|y| \geq 1$ by assumption. This establishes (i) and (ii), and completes this part of the proof.

We have now shown that $a = a'$, $c = c'$, so it remains to show that $b = b'$. That $d = d'$ immediately implies that $b = \pm b'$. Since $b \geq 0$ if $a < c$, and similarly for b', we are already done in this case. So we may assume that $a < c$. Also we cannot have b or b' equalling $-a$, so we may also assume that $|b| = |b'| < a$. But then the sequence of inequalities above implies that $(x, y) = (\pm 1, 0)$ are the only proper representations of a by f, and that $(x, y) = (0, \pm 1)$ are the only proper representations of c by f. Let now $M = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ be an element of $\mathrm{SL}_2(\mathbf{Z})$ which transforms f to f'. The equations (66) imply that

$$\begin{aligned} a' = a &= f(p, r), \\ c' = c &= f(q, s). \end{aligned}$$

But $\det(M) = 1$ implies that $\mathrm{GCD}(p, r) = \mathrm{GCD}(q, s) = 1$. Hence we must have $(p, r) = (\pm 1, 0)$ and $(q, s) = (0, \pm 1)$. Finally, then, $\det(M) = 1$ implies that $M = \pm I$ and hence transforms any form to itself. Thus $f = f'$ and we are done.

Important remark 1 : The proof of part (i) gives an algorithm for producing a reduced form equivalent to a given form (in class I considered the example of the form $f(x, y) = 22x^2 - 108xy + 133y^2$, which I reduced to the form $f'(x, y) = 2x^2 + 5y^2$). Because of part (iii) this gives, in the case of positive definite forms, an algorithm for deciding whether two given forms are equivalent.

However, for indefinite forms, things get much more complicated. It can happen that a single equivalence class of indefinite forms contains more than one reduced member. In particular, the above algorithm for deciding equivalence of forms does not work in general. In fact, one needs a whole new (and rather different !) reduction theory in order to be able to find such an algorithm[4].

---

[4] We will not have time to describe this theory, but here are the main results. For detailed proofs, see my handout from Zagier's book, 'Zetafunktionen und quadratische Körper'.

In this new theory, a form $\{a, b, c\}$ is called *reduced* if $a > 0, c > 0$ and $a + c < b$. We consider a special type of transformation (65). If f is the form $\{a, b, c\}$, set $n = \lceil \frac{b+\sqrt{d}}{2a} \rceil$ and define $T(f)$ to be the form obtained by transforming, as in (65), with the matrix

IMPORTANT REMARK 2 : Motivated by Theorem 17, we define $H(d)$ to be the number of equivalence classes of binary forms of discriminant d. The theorem says that $H(d)$ is finite, and because of the principal forms, we conclude that $H(d)$ is a non-zero positive integer. For positive definite forms, part (iii) of the theorem implies that, in order to compute $H(d)$, it suffices to compute the number of reduced forms of discriminant d. This can be done with the help of the inequality (70). For indefinite forms, the analogous procedure using (71) only gives, in general, an upper bound for $H(d)$. One of the classical results of algebraic number theory is

Theorem 18 (Baker/Stark 1966)[5] *There are only finitely many $d < 0$ for which $H(d) = 1$, and these are given explicitly by*
$d = -3, -4, -7, -8, -11, -19, -43, -67, -163.$

Likewise, one of the outstanding open problems in algebraic number theory is

Open problem *Do there exist infinitely many $d > 0$ for which*
$H(d) = 1$ *?*

This completes our discussion of the reduction theory for binary quadratic forms. We now continue our prepartions for a discussion of the representation problem.

$M_n = \begin{pmatrix} n & 1 \\ -1 & 0 \end{pmatrix}$. Then the main result of the theory is

Theorem. (i) The number of reduced forms of a given discriminant is finite.
(ii) For any given form f, there exists a positive integer k, depending on f, such that $T^k(f)$ is reduced.
(iii) The operator T maps reduced forms to reduced forms.
(iv) Every equivalence class of reduced forms is acted upon transitively by T.

Note that the theorem does not say that each class of forms contains a unique reduced form (this is false !), hence does not provide an algorithm for computing class numbers (see important remark no. 2 above). On the other hand, it is easy to see that (ii),(iii) and (iv) give an algorithm for deciding equivalence of any 2 given forms.

[5] This result is usually phrased in terms of which imaginary quadratic number fields have unique factorisation. We will explain what this means later. Likewise, the open problem following Theorem 18 is usually phrased in terms of unique factorisation for real quadratic fields.

Definition : En matris $M \in SL_2(\mathbf{Z})$ sägs vara en *automorfism* av formen f om

$$M^T A_f M = A_f. \tag{72}$$

Notera att automorfismerna av f utgör en delGRUPP av $\mathrm{SL}_2(\mathbf{Z})$. Denna grupp betecknas $\mathrm{Aut}(f)$. Ordningen av gruppen betecknas med $w(f)$ om f är definit. Vi kommer att visa såsmåningom att en indefinit form har oändligt många automorfismer.

Proposition 16 *(i) Om $X \in \mathbf{Z}^2$ och $M \in Aut(f)$, då är*

$$f(MX) = f(X). \tag{73}$$

(ii) Om $f \sim g$, och P är en matris så att $P^T A_f P = A_g$, då har vi en 1-1 korrespondens

$$\begin{array}{ccc} Aut(f) & \leftrightarrow & Aut(g) \\ M & \leftrightarrow & P^{-1}MP. \end{array} \tag{74}$$

Notera att del (i) av propositionen säger att när vi räknar antalet representationer av ett heltal n med f, så måste vi inte glömma automorfiska representationer.

Del (ii) av propositonen säger att ekvivalenta former har samma antalet automorfismer.

Det viktiga begreppet för att studera automorfism frågan är

Definition : En form $f = \{a, b, c\}$ kallas för *primitiv* om $\gcd(a, b, c) = 1$.

Proposition 17 *Låt $f = \{a, b, c\}$ vara en form med $gcd(a, b, c) = r$. Då är formen $f' = \{a/r, b/r, c/r\}$ primitiv och $Aut(f) = Aut(f')$.*

Bevis : Klart.

Denna proposition reducerar beräkningen av automorfismgrupper till den för primitiva former. Det visar sig nu att automorfismgrupperna av primitiva former kan beskrivas helt explicit. Vi har

Sats 19 *Låt $f = \{a, b, c\}$ vara en primitiv form av diskriminant d. Då består*

$Aut(f)$ precist av alla matriser

$$M = \begin{pmatrix} \frac{1}{2}(t-bu) & -cu \\ au & \frac{1}{2}(t+bu) \end{pmatrix}, \tag{75}$$

där $(t,u) \in \mathbf{Z}^2$ är en lösning av ekvationen

$$t^2 - du^2 = 4. \tag{76}$$

Bevis : Satz 202 i Landaus bok (se utdelad stencil).

Ekvation (76) kallas för *Pells ekvation.* I fallet $d < 0$ kan lösningarna ses direkt, och de är

$$\begin{array}{rc} (\pm 2,0),(0,\pm 1), & \text{om } d=-4, \\ (\pm 2,0),\pm(1,1),\pm(1,-1), & \text{om } d=-3, \\ (\pm 2,0), & \text{annars.} \end{array} \tag{77}$$

Då följer det från Sats 19 att

Korollarium 20 *Låt $f = \{a,b,c\}$ vara en primitiv positiv definit form av diskriminant $d < 0$. Då har vi att*

$$\begin{array}{rlc} Aut(f) = \{\pm M_1, \pm M_2\}, & w(f)=4, & \textit{om } d=-4, \\ Aut(f) = \{\pm M_1, \pm M_3, \pm M_3^2\}, & w(f)=6, & \textit{om } d=-3, \\ Aut(f) = \{\pm M_1\}, & w(f)=2, & \textit{annars,} \end{array} \tag{78}$$

där

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}. \tag{79}$$

Speciellt så är $Aut(f)$ alltid en ändlig cyklisk grupp.

För indefinita former, är läget mycket mer interessant. Man får följande resultat

Sats 21 *(i) (*Lagrange 1768*) Låt d vara ett positivt heltal som inte är en perfekt kvadrat. Då har Pells ekvation (76) oändligt många lösningar. Om (t_0,u_0) är den lösning med $t_0,u_0 > 0$ och u_0 minimal, då ges alla lösningar av*

$$\frac{1}{2}(t+u\sqrt{d}) = \pm\left[\frac{1}{2}(t_0+u_0\sqrt{d})\right]^m, \tag{80}$$

där $m \in \mathbf{Z}$.

*(ii) (*Gauss*) Om* f *är en primitiv form av diskriminant* d*, då är* $Aut(f) \cong \mathbf{Z}$*, en oändlig cyklisk grupp.*

Bevis : Kommer senare. Nuförtiden, genomförs beviset bäst i sättningen av algebraiska talkropper. Naturligtvis kan man undvika denna teori, som Lagrange tvingades göra - för en sådan synvinkel, se t.ex. Landaus Vorlesungen, Satz 111.

Lecture 15 : 061204

Vi vill säga lite mer om primitiva former innan vi går vidare.

Proposition 18 *Om f är primitiv och $f \sim g$, då är g också primitiv.*

DEFINITION : Vi betecknar med $h(d)$ antalet ekvivalens klasser av primitiva former av diskriminant d. Notera att detta är väl-definierad enligt Proposition 18.

DEFINITION : Ett heltal $d \equiv 0$ eller 1 (mod 4) kallas för en *fundamental diskriminant* om varje form av diskriminant d är primitiv.

Proposition 19 *Låt $d \equiv 0$ eller 1 (mod 4).*
(i) $H(d) = h(d) \Leftrightarrow d$ är en fundamental diskriminant.
(ii)

$$H(d) = \sum_{t^2 | d} h\left(\frac{d}{t^2}\right). \tag{81}$$

(iii) d är en fundamental diskriminant omm

$$\begin{aligned} d \text{ är kvadratfri}, &\quad \text{om } d \equiv 1 \pmod 4, \\ d/4 \text{ är kvadratfri och } \equiv 2 \text{ eller } 3 \pmod 4, &\quad \text{om } d \equiv 0 \pmod 4. \end{aligned} \tag{82}$$

BEVIS : Enkelt och tråkigt - lämnas till läsaren.

Representationsproblemet

We now finally come to what we've really been building up to all this time, namely the analysis of the equation

$$f(x,y) = ax^2 + bxy + cy^2 = n, \tag{83}$$

where everything in sight is an integer. Any such problem has both a theoretical and a practical aspect. From the theoretical viewpoint, one seeks some 'nice formulas' for the numbers $R(n,f)$ of solutions to (83). From the practical viewpoint, one is interested in having fast algorithms for computing all solutions explicitly[6]. Obviously, these two issues are inextricably linked.

[6]The results we present in this course will only be to the extent of showing that algorithms exist. I have no idea what the state of the art is in regard to 'fast' algorithms, and hence will make no attempt to go into this matter.

Keeping in mind what people have actually succeeded in doing with the problem, we will follow the following program :

STEP 1 : As we've described earlier, the problem is uninteresting when the discriminant d is a perfect square. So we'll alawys assume this is not the case. Also, we lose nothing, in the case $d < 0$, by assuming that f is positive definite, i.e.: that $a, c > 0$.

STEP 2 : Right at the outest, a clear distinction presents itself between the definite and indefinite cases. In the former case, the number $R(n, f)$ is always finite, and the coefficients (x, y) of a representation are explicitly bounded by

$$|x| \leq \sqrt{4nc/|d|}, \qquad |y| \leq \sqrt{4na/|d|}. \tag{84}$$

These bounds follow easily from completion of squares, as in (64). Note that we hereby immediately obtain a (slow) algorithm for finding all solutions of (83) in the definite case - just check all pairs x, y up to the bounds in (84).

In the indefinite case, it follows immediately from Sats 21 that $R(n, f)$ is always either 0 or ∞. That is, each solution gives rise to an infinity of others by the application of the automorphisms of the form (see Prop. 16). It turns out that there is a fairly natural way of selecting, from each such infinite set of automorphic representations, one which is called *primary*. The number of primary representations then turns out to be finite, and we adopt the notation $R(n, f)$ for this number instead. The remarkable and wonderful fact is that, with this new convention, the indefinite and definite cases become 'unified' : more precisely we get explicit formulas valid in both cases (see Step 3).

With regard to the algorithmic question of finding explicit (primary) representations, it should not surprise you that this reduces to finding generating solutions of Pell's equation. This is an instance of a problem in *Diophantine approximation* theory, and so we'll be lead to roam into this area. However, this will be postponed until nearer the end of the course.

STEP 3 : It is convenient to first of all count proper representations. The number of proper representations of n by f is denoted $r(n, f)$, with the extra condition that the representations be primary in the indefinite case.

Observe that

$$R(n,f) = \sum_{t^2|n} r\left(\frac{n}{t^2}, f\right). \tag{85}$$

We now come to the crucial point :

As indicated in Step 2, there are algorithms for computing the numbers $r(n,f)$, for any given n and f. However, there seems to be NO SIMPLE FORMULAS for them. But if we sum over the forms of a fixed discriminant, then we get REALLY NICE formulas. More precisely, define, for each discriminant d,

$$R_d(n) := \sum_f R(n,f), \qquad r_d(n) := \sum_f r(n,f), \tag{86}$$

where the sum runs over a representative set of forms of discriminant d, one from each equivalence class. By Prop. 15, the sums are then well-defined.

The number $r_d(n)$ can be expressed in terms of the number of solutions to a certain quadratic congruence (Theorems 22,23 and 24 below). But solving quadratic congruences is a relatively simple matter, as discussed earlier in this course (see, in particular, Prop. 9). In particular, this already takes care of the algorithmic issues. The resulting 'formulas' are simplest in the case when d is a fundamental discriminant (basically, this is because all forms of discriminant d then have the same number of automorphisms - see Korollarium 20) and $(n,d) = 1$.

Representationsproblemet : the results !!

We start with acouple of lemmas, which will be used in what follows :

Lemma 4 *Let x, y be integers such that $SGD(x,y) = 1$. Then there are infinitely many matrics in $SL_2(\mathbf{Z})$ whose first column is $\begin{pmatrix} x \\ y \end{pmatrix}$. If $M_0 = \begin{pmatrix} x & u \\ y & v \end{pmatrix}$ is any such matrix, then the complete list of such matrices is given by*

$$\left\{ M_t := \begin{pmatrix} x & u+tx \\ y & v+ty \end{pmatrix}, \ t \in \mathbf{Z} \right\}.$$

Proof : Follows easily from Euclid's lemma.

Lemma 5 *Låt f, g vara ekvivalenta binära former. Då finns det precist $w(f)$ matriser $M \in SL_2(\mathbf{Z})$ så att $M^T A_f M = A_g$.*

Bevis : Eftersom $f \sim g$, finns det minst en matris N så att $N^T A_f N = A_g$. Om N_1, N_2 är två sådana matriser, sätt $M = N_1 N_2^{-1}$ och kontrollera att $M^T A_f M = A_f$, dvs att $M \in \mathrm{Aut}(f)$. Därför, $\{MN : M \in \mathrm{Aut}(f)\}$ är mängden av alla matriser som tar f till g. Denna mängd har $w(f)$ element.

We are now ready to give the basic connection between proper representations and quadratic congruences :

Theorem 22 *The number n is properly represented by some binary quadratic form of discriminant d iff the congruence*

$$h^2 \equiv d \ (mod\ 4n) \tag{87}$$

is solvable.

Proof : (As in Baker, p.37-8). First suppose the congruence is solvable and let h be any solution. Then there exists an integer k such that $h^2 = d + 4nk$, hence

$$h^2 - 4nk = d. \tag{88}$$

Consider now the quadratic form

$$f(x, y) = nx^2 + hxy + ky^2.$$

Then f has discriminant d, by (88). But $f(\pm 1, 0) = n$ so f properly represents n.

Conversely, suppose n is properly represented by some form f of discriminant d, say $f(x, y) = d$, where $\mathrm{GCD}(x, y) = 1$. Let M be any matrix in $\mathrm{SL}_2(\mathbf{Z})$ whose first column is $\begin{pmatrix} x \\ y \end{pmatrix}$: such a matrix exists by Lemma 4. Consider the form $g = \{a, b, c\}$ given by

$$A_g = M^T A_f M.$$

Then $f \sim g$ so g also has discriminant d. By eqs. (66) we have that $a = f(x, y) = n$. Hence $d = d_g = b^2 - 4ac = b^2 - 4nc$, which implies that $b^2 \equiv d \ (\mathrm{mod}\ 4n)$. Hence (87) has a solution, v.s.v.

Lektion 16 : 081204

Vi börjar med en tillämpning av Sats 22.

Proposition 20 *Låt p vara ett udda primtal. Då finns det en form av diskriminant d som representerar p omm antingen $p|d$ eller* $\left(\frac{d}{p}\right) = 1$.

Bevis : Enligt Sats 22, finns det en form av diskriminant d som representerar p omm kongruensen $h^2 \equiv d \pmod{4p}$ är lösbar. Eftersom $d \equiv 0$ eller $1 \pmod 4$ så är $h^2 \equiv d \pmod 4$ lösbar. Därmed är $h^2 \equiv d \pmod{4p}$ lösbar, enligt Proposition 9, om och endast om $h^2 \equiv d \pmod p$ är det ; m.a.o. om och endast om antingen $p \mid d$ eller $\left(\frac{d}{p}\right) = 1$.

Exempel : Låt p vara ett udda primtal. Då har ekvationen $x^2 - 2y^2 = p$ en lösning $(x, y) \in \mathbf{Z}^2$ omm $p \equiv \pm 1 \pmod 8$.

För formen $x^2 - 2y^2$ har diskriminant 8 och m.h.a. (71) visar man lätt att den är den unika reducerade formen av diskriminant 8. Enligt Sats 17 finns det då precis en klass av former av diskriminant 8, så att i detta fall medför Sats 22 och Proposition 20 att $x^2 - 2y^2$ representerar udda primtalet p omm antingen $p \mid 8$, som är omöjligt, eller $\left(\frac{8}{p}\right) = 1 \Leftrightarrow \left(\frac{2}{p}\right) = 1 \Leftrightarrow p \equiv \pm 1 \pmod 8$, enligt Proposition 11.

Under resten av lektionen, koncentrerar vi på representationsproblemet för positiv definita former. Vi skall indikera hur resultaten kan utvidgas till det indefinita fallet (där man betraktar bara primära representationer) och hänvisa den intresserade läsaren till Landaus Vorlesungen för bevis, och till och med för definitionen av en primär representation, som är lite krånglig.

Det första målet är att ge en mer precis version av Sats 22.

Notation : Låt $d < 0, n > 0$. Låt $h \in \mathbf{Z}$ satisfiera $h^2 \equiv d \pmod{4n}$, säg $h^2 = d + 4nk$. Den kvadratiska formen $nx^2 + hxy + ky^2$ ska betcknas med $f_{n,h}$. Notera att $d(f_{n,h}) = d$.

Sats 23 *Låt $f = \{a, b, c\}$ vara en positiv definit form av diskriminant $d < 0$. Låt $n > 0$. Sätt*

$$H_f(n) \stackrel{def}{=} \#\{h : 0 \leq h < 2n,\ h^2 \equiv d \ (mod\ 4n),\ f_{n,h} \sim f\}. \tag{89}$$

Då är

$$r(n,f) = w(f)H_f(n). \tag{90}$$

Bevis : Låt X vara mängden av alla former $f_{n,h}$ så att h satisfierar villkoren i definitionen av $H_f(n)$. Denna mängd innehåller då $H_f(n)$ former. Enligt Lemma 5 finns det, för varje form $f_{n,h} \in X$, precis $w(f)$ matriser $M \in SL_2(\mathbf{Z})$ som tar f till $f_{n,h}$. Då finns det totalt $w(f)H_f(n)$ matriser som tar f till någon form i X. Låt mängden av dessa matriser betecknas med Y. För att bevisa (90), räcker det nu att etablera an 1-1 korrespondens mellan matriserna i Y och egentliga representationer av n med f.

$\Rightarrow$ Först, välj $M \in Y$, säg $M = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$. Låt $M(f) = f_{n,h} = nx^2 + hxy + ky^2$. Från transformationsformlerna (66) ser vi att

$$f(\alpha,\beta) = n. \tag{91}$$

Eftersom $\det(M) = 1$ måste $\gcd(\alpha,\beta) = 1$, så att (91) är en egentlig representation av n med f.

$\Leftarrow$ Vi måste visa att ovanstående korrespondensen har en invers, dvs vi måste bevisa följande :

Påstående : Låt $f(x,y) = n$ vara en egentlig representation av n med f. Då finns det en unik matris $M = \begin{pmatrix} x & * \\ y & * \end{pmatrix} \in SL_2(\mathbf{Z})$ så att $M \in Y$.

För att bevisa detta, så använder vi Lemma 4, som säger att alla matriser i $SL_2(\mathbf{Z})$ med första kolonn $\begin{pmatrix} x \\ y \end{pmatrix}$ ges av $M_t = \begin{pmatrix} x & v+tx \\ y & u+ty \end{pmatrix}$, där $t \in \mathbf{Z}$ och $xu - yv = 1$.

Nu säger påståendet att det finns precist ett t så att $M_t \in Y$.

Vi har redan sett i beviset av Sats 22 i måndags att varje matris M_t tar f till en form av typen $nx^2 + hxy + ky^2$, där $h^2 \equiv d \pmod{4n}$, och $h^2 = d + 4nk$. Målet nu är att bevisa att det finns ett unikt t så att denna h också satisfierar villkoret $0 \leq h < 2n$.

Från transformationsformlerna (66) får vi att

$$\begin{array}{rcccc} h & = 2ax(v+tx) + b[x(u+ty)+y(v+tx)] + 2cy(u+ty) \\ & = & 2t(ax^2+bxy+cy^2) & + & (2axv+bxu+byv+2cyu) \\ & = & 2tf(x,y) & + & K \\ & = & 2nt & + & K, \end{array}$$

där konstanten K är oberoende av t. Då ser vi att det finns ett unikt t så att $0 \leq h < 2n$, q.e.d.

Anmärkning 1 : Om f är en primitiv, indefinit form av diskriminant $d > 0$, då stämmer (90) fortfarande om vi tar $w(f) = 1$. Se Vorlesungen, Satz 203, för ett bevis.

Anmärkning 2 : Ekv. (90) ger en ny algoritm för att beräkna $r(n, f)$. Först, beräknar man $w(f)$, som är lätt pga Prop. 17 och Korollarium 20. Då lösar man kongruensen $h^2 \equiv d \pmod{4n}$, och för varje lösning kontrollerar om $f_{n,h} \sim f$. Man använder reduktionsteorin för denna sista del. Notera att samma algoritm funkar i det indefinita fallet, även om reduktionsteorin är ganska mer komplicerad.

Det följer direkt från (90) att

$$r_d(n) = \sum w(f) H_f(n), \tag{92}$$

där summan är över en mängd av representanter för ekvivalensklasserna av former av diskriminant d. Eftersom $\sum H_f(n)$ är antalet lösningar till en kvadratisk kongruens, misstänker man att det borde finnas en snygg formel för $r_d(n)$ om $w(f)$ är en konstant i summan (92). Från Korollarium 20 ser vi att det är så, t.ex. när d är en fundamental diskriminant[7]. Då har vi bevisat

Sats 24 *Låt $d < 0$ vara en fundamental diskriminant. Då är*

$$r_d(n) = wN_d(n), \tag{93}$$

där

$$w = \begin{cases} 2, & \text{om } d < -4, \\ 4, & \text{om } d = -4, \\ 6, & \text{om } d = -3, \end{cases} \tag{94}$$

[7] För en komplett karakterisering av vilka d satisfierar denna egenskap, se supplementär övning nr. 13.

och

$$N_d(n) = \#\{h : h^2 \equiv d\ (mod\ 4n),\ 0 \leq h < 2n\}. \tag{95}$$

Anmärkning : Resultatet gäller fortfarande för $d > 0$, om vi tar $w = 1$.

Lecture 17 : 101204

It is beyond the scope of this course to present any general theory for quadratic forms in more than two variables. We content ourself with presenting, in addition to the earlier results of Pythagoras and Fermat, a single famous result :

Theorem 25 (Lagrange 1770) *Every positive integer is the sum of four squares.*

PROOF : (Following Baker, p.39-40). First, because of the identity (see the remark below for a discussion of where this comes from)

$$\begin{aligned} &(x^2+y^2+z^2+w^2)(a^2+b^2+c^2+d^2) \\ &= (xa+yb+zc+wd)^2 + (xb-ya+wc-zd)^2 \\ &+(xc-za+yd-wb)^2 + (xd-wa+zb-yc)^2, \end{aligned} \tag{96}$$

it suffices to prove the result for primes. And since $2 = 1^2+1^2+0^2+0^2$, we may confine ourselves to odd primes.

So let p be an odd prime. Let $l > 0$ be such that lp is the smallest non-zero multiple of p expressible as the sum of four squares. Our aim is to show that $l = 1$. We achieve this in several steps :

Step 1 : $l < p$.

As x runs over all residue classes modulo p, so x^2 runs over $\frac{p+1}{2}$ distinct classes. Similarly, as y runs over all classes mod p, so $-1-y^2$ runs over $\frac{p+1}{2}$ different classes. By the Pigeonhole principle, and the fact that $a^2 \equiv (-a)^2 \pmod{p}$, there exist $x, y \in [0, p/2)$ such that $x^2 \equiv -1-y^2 \pmod{p}$, hence $x^2+y^2+1 = rp$, for some integer r. Thus rp is a sum of four squares. But $x, y \in [0, p/2) \Rightarrow r < p$. Hence $l < p$, as required.

Step 2 : l is odd.

Suppose

$$x^2+y^2+z^2+w^2 = rp,$$

where r is even. Then rp is even, hence an even number of x, y, z and w have to be even. Hence, WLOG, $x \equiv y \pmod 2$ and $z \equiv w \pmod 2$. But then

$$\left(\frac{r}{2}\right) p = \left(\frac{x+y}{2}\right)^2 + \left(\frac{x-y}{2}\right)^2 + \left(\frac{z+w}{2}\right)^2 + \left(\frac{z-w}{2}\right)^2,$$

and the HL is a sum of four integer squares. This proves that l must be odd.

Step 3 : We now suppose that $l > 1$ and obtain a contradiction. Let

$$x^2 + y^2 + z^2 + w^2 = lp \tag{97}$$

be any representation of lp as a sum of four integer squares. Let a, b, c, d be the numerically least residues of x, y, z and w respectively, modulo l, as defined in the statement of Gauss' lemma. Then $a^2 + b^2 + c^2 + d^2 \equiv x^2 + y^2 + z^2 + w^2 \equiv 0 \pmod l$, say

$$a^2 + b^2 + c^2 + d^2 = kl. \tag{98}$$

Since l is odd, each of a, b, c and d lies in the OPEN interval $(-l/2, l/2)$ and hence $k < l$. By (96), the number $(kl)(lp) = l^2(kp)$ can be written as the sum of four integer squares, which we denote E, F, G and H. By inspection of (96) and the fact that $x \equiv a, y \equiv b, z \equiv c$ and $w \equiv d \pmod l$, we see that each of E, F, G and H is divisible by l. Hence, dividing across by l^2, we find that

$$kp = \left(\frac{E}{l}\right)^2 + \left(\frac{F}{l}\right)^2 + \left(\frac{G}{l}\right)^2 + \left(\frac{H}{l}\right)^2,$$

is a sum of four integer squares. Since $k < l$, this contradicts the definition of l unless $k = 0$. But if $k = 0$ then, by (98), each of a, b, c and d equals zero, hence each of x, y, z and w is divisible by l. But then the VL of (97) is divisible by l^2, which implies that $l \mid p$. But p is a prime so either $l = p$, which is impossible by *Step 1*, or $l = 1$, v.s.v.

Remark 1 As with most (all ?) proofs of this theorem, Baker begins by reducing to primes via the identity (96). To understand where this identity really comes from, one should look at Hamilton's *quaternions* H. Recall that these are defined by

$$H := \{\mathbf{R} + \mathbf{R}i + \mathbf{R}j + \mathbf{R}k : i^2 = j^2 = k^2 = -1, ij = k, jk = i, ki = j\}. \tag{99}$$

H is thus a vector space of dimension 2 over $\mathbf{C}$ (or 4 over $\mathbf{R}$), which can be proven to be a (non-commutative) division algebra. A classic theorem of Frobenius ([1], Theorem 7.3.1) states that every division ring, algebraic over $\mathbf{R}$, is isomorphic to $\mathbf{R}$, $\mathbf{C}$ or H.

Let $\alpha = x + yi + zj + wk := (x, y, z, w) \in H$. The *conjugate* of α, denoted α^*, is defined by $\alpha^* := (x, -y, -z, -w)$. The following properties are easily verified :

$$[\alpha^*]^* = \alpha \tag{100}$$

$$\alpha\alpha^* = x^2 + y^2 + z^2 + w^2 \in \mathbf{R} \tag{101}$$

$$(\alpha\beta)^* = \beta^*\alpha^*. \tag{102}$$

Because of (100) we have a map $N : H \to \mathbf{R}$, called the *norm* map, given by $N(\alpha) := \alpha\alpha^*$. Eqs. (99) - (101) then easily imply that

$$N(\alpha\beta) = N(\alpha)N(\beta), \tag{103}$$

which is equivalent to the identity (96).

Remark 2 Lagrange's theorem says that every integer is expressible as the sum of four squares. Theorem 8 easily implies that the integer n is expressible as the sum of two squares if and only if every prime $\equiv 3 \pmod 4$ appears to an even power in the prime factorisation of n. It therefore remains to decide which integers are sums of three squares. This issue was settled by Gauss :

Theorem 26 (Gauss) *The positive integer n is a sum of three squares if and only if it is not of the form $4^j(8l + 7)$, for some $j, k \geq 0$.*

Proof (optional) : [2], Chapter IV, Appendix.

At about the same time that Lagrange proved his theorem, Waring proposed the following far-reaching generalisation of it :

Waring's conjecture *For every integer $k > 0$, there exists an integer $G(k)$, depending only on k, such that every positive integer is the sum of at most $G(k)$ k:th powers of positive integers.*

This conjecture was first proven by Hilbert in 1909, by means of a highly complicated combinatorial argument. A more elegant (and more widely applicable !) analytic treatment was given by Hardy and Littlewood in the 1920s. For an introduction to what has since become known as the *Hardy-Littlewood circle method*, see [3].

References

[1] I.N. Herstein, Topics in Algebra, Wiley.
[2] J-P. Serre, A Course in Arithmetic, Springer (GTM No. 7).
[3] R.C. Vaughan, The Hardy-Littlewood Method, Cambridge University Press.

Lectures 18-19 : 13-151204

These two lectures comprise a very short introduction to analytical methods in number theory. Our very modest goals will be to define the (Riemann) zeta-function, explain Euler's method of proving that the sum of the reciprocals of the primes diverges, and indicate how Dirichlet adapted this method to prove his celebrated theorem on primes in arithmetic progressions. We will not have time, however, to give the full proof of Dirichlet's theorem. For that, the interested reader is referred to my 2000 lecture notes and to Davenport's book.

NOTATION : Complex numbers will be denoted by s. We write $s = \sigma + it$, so that $\mathrm{Re}(s) = \sigma$ and $\mathrm{Im}(s) = t$.

DEFINITION : Let $s \in \mathbf{C}$ with $\sigma > 1$. The *Riemann zeta-function* is given by

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

It is clear that the sum converges absolutely when $\sigma > 1$ and uniformly in any strip $\sigma > 1 + \delta$. Hence by Weierstrass' theorem, it defines an analytic function in the half-plane $\sigma > 1$. If any of this sounds like double-dutch to you, then I suggest you revise your complex analysis a little. However, you won't lose much by just accepting it as fact.

Proposition 21 (Euler) *Let $\sigma > 1$. Then*

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}. \tag{104}$$

In other words, the infinite product converges absolutely in this region and coincides with $\zeta(s)$.

PROOF : We did not give a rigorous proof (neither did Euler, probably), but just the idea. One way of proving (104) is to show that, for every integer $N > 0$,

$$\left[\prod_{p \leq N} \left(1 - \frac{1}{p^s}\right)\right] \cdot \zeta(s) = \sum_{p \nmid n \text{ for any } p \leq N} \frac{1}{n^s}.$$

As $N \to \infty$ the HL will converge to 1, implying (104).

An alternative is to start with the infinite product and note that, by the binomial theorem, each factor can be expanded as an infinite series

$$\left(1 - \frac{1}{p^s}\right)^{-1} = \sum_{m=0}^{\infty} \frac{1}{p^{ms}}.$$

Thus the infinite product becomes

$$\prod_p \left(\sum_{m=0}^{\infty} \frac{1}{p^{ms}} \right),$$

and when one multiplies this out, one sees that, for every $n > 0$, the term $1/n^s$ appears exactly once, by FTA.

TERMINOLOGY : An expression of the form

$$\prod_p f_p(p^s),$$

where each $f_p(x)$ is a rational function, is called an *Euler product*.

Euler used his product representation for ζ to prove a stronger form of Euclid's result that there are infinitely many primes :

Theorem 27 (Euler) *There are infinitely many primes and the sum of their reciprocals diverges.*

PROOF : Taking log of both sides of (104)[8] we get, for $\mathrm{Re}(s) > 1$,

$$\log \zeta(s) = -\sum_p \log \left(1 - \frac{1}{p^s}\right). \tag{105}$$

Next recall that the Taylor series for the log function, valid when $|z| < 1$, is given by

$$-\log(1 - z) = \sum_{m=1}^{\infty} \frac{1}{m} z^m. \tag{106}$$

[8] At all times, unless otherwise stated, we are using the principal branch of the log function.

Substituting (106) into (105) we get, also for $\mathrm{Re}(s) > 1$,

$$\log \zeta(s) = \sum_p \sum_{m=1}^{\infty} \frac{1}{mp^{ms}}. \tag{107}$$

Now group the terms on the HL into two groups, those with $m = 1$ and those with $m > 1$. Note that we are changing the order of summation here, but that is okay because the series is absolutely convergent when $\mathrm{Re}(s) > 1$[9]. We obtain

$$\log \zeta(s) = \sum_p p^{-s} + \sum_p \sum_{m=2}^{\infty} \frac{1}{mp^{ms}}. \tag{108}$$

The idea now is to look at what happens when $s \to 1^+$. The sum over $m \geq 2$ does not blow up - in fact, it's value at $s = 1$ can easily be shown to be less than, for example, $2\zeta(2) = \pi^2/3$. On the other hand, we all know from envariabelanalys (use, say, the integral test) that $\zeta(s) \to +\infty$ as $s \to 1^+$. This means that the VL of (108) goes to infinity as $s \to 1^+$. It follows that the same is true of the rhs, and hence that

$$\lim_{s \to 1^+} \sum_p p^{-s} = \infty. \tag{109}$$

In particular, the sum must contain infinitely many terms (i.e.: there are infinitely many primes) and $\sum p^{-1}$ diverges. This proves Theorem 27.

It is a famous theorem of Riemann that $\zeta(s)$ has a meromorphic continuation to the whole complex plane, with a single simple pole at $s = 1$. It is beyond the scope of this course to prove this result, however we would like to prove something weaker. For this we will use a trick called *Abel summation*, which will also be used again later. We state it explicitly for convenience :

Abel summation formula *Let* $(a_n)_1^{\infty}$ *and* $(b_n)_1^{\infty}$ *be any two sequences of complex numbers. For eacn* $n > 0$ *denote* $A_n := \sum_{i=1}^{n} a_i$*. Then for any* $N > 0$*, we have*

$$\sum_{n=1}^{\infty} a_n b_n = \sum_{n=1}^{N} A_n (b_n - b_{n+1}) + A_{n+1} b_{n+1} + \sum_{n=N+2}^{\infty} a_n b_n. \tag{110}$$

[9] Recall from envariabelanalys (or see p.42 of my notes on flervariabelanalys from 1999) that the sum of an absolutely convergent series is independent of the ordering of the terms. This is not true for conditionally convergent series (Riemann's theorem).

In particular, under suitable conditions of convergence, we have

$$\sum_{n=1}^{\infty} a_n b_n = \sum_{n=1}^{\infty} A_n (b_n - b_{n+1}). \tag{111}$$

Now we shall prove

Theorem 28 *The zeta-function has a meromorphic continuation to the half-plane $\sigma > 0$, with a single simple pole at $s = 1$.*

Sketch Proof : Using (111), with $a_n = 1, b_n = 1/n^s$, we can rewrite the sum for $\zeta(s)$, in the region $\sigma > 1$, as

$$\zeta(s) = \sum_{n=1}^{\infty} n[n^{-s} - (n+1)^{-s}].$$

We note that, for any $a, b \in \mathbf{R}$,

$$\int_a^b \frac{1}{x^{s+1}}\, ds = \frac{1}{s}\left(a^{-s} - b^{-s}\right).$$

Hence, for $\sigma > 1$,

$$\begin{aligned}\zeta(s) &= s \cdot \left(\sum_{n=1}^{\infty} n \int_n^{n+1} \frac{1}{x^{s+1}}\, ds \right) \\ &= s \cdot \sum_{n=1}^{\infty} \int_n^{n+1} \frac{[x]}{x^{s+1}}\, ds \\ &= s \int_1^{\infty} \frac{[x]}{x^{s+1}}\, ds,\end{aligned}$$

where $[\cdot]$ denotes 'integer part of'. For any real number x, let us denote $(x) := x - [x]$. Then we may continue

$$\begin{aligned}\zeta(s) &= s \int_1^{\infty} \frac{x - (x)}{x^{s+1}}\, ds \\ &= s \left[\int_1^{\infty} \frac{1}{x^s}\, ds - \int_1^{\infty} \frac{(x)}{x^{s+1}}\, ds \right] \\ &= s \left[\frac{1}{s-1} - \int_1^{\infty} \frac{(x)}{x^{s+1}}\, ds \right].\end{aligned}$$

In summary, we have shown that, when $\sigma > 1$,

$$\zeta(s) = \frac{s}{s-1} - s \int_1^\infty \frac{(x)}{x^{s+1}} \, ds. \tag{112}$$

But, since $(x) \in [0, 1)$ for any x, the integral in (112) is absolutely convergent in the region $\sigma > 0$, hence defines an analytic function of s in that region. Thus (112) is the desired meromorphic continuation of ζ, and note that the leading term on the HL contributes a single simple pole at $s = 1$.

We now turn to Dirichlet's generalisation of Euler's method.

DEFINITION : Let $(a_n)_1^\infty$ be any sequence of complex numbers. A function $D(s)$ defined by a series of the form

$$D(s) := \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

is called a *Dirichlet series*.

Proposition 21 has the following generalisation

Proposition 22 *Låt $(a_n)_1^\infty$ vara en följd av komplexa tal så att*

$$a_1 = 1 \quad \textit{och} \quad a_{mn} = a_m a_n \;\; \textit{för alla} \; m, n \geq 1. \tag{113}$$

Då är

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s} = \prod_p \left(1 - \frac{a_p}{p^s}\right)^{-1}, \tag{114}$$

när som helst båda sidorna konvergerar och definierar analytiska funktioner.

BEVIS : Det är exakt samma idé som i beviset av Proposition 21.

In order to prove his theorem, Dirichlet studied a particular class of Dirichlet series :

DEFINITION : En Dirichlet serie på formen

$$L(s, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}, \tag{115}$$

där χ är en karaktär modulo q, för något heltal q, kallas för en *Dirichlet L-serie*[10].

Notera att om $q = 1$ då måste $\chi = \chi_0$ vara den triviala karaktären och $L(s, \chi) = \zeta(s)$. Mer allmänt, om $\chi = \chi_0$ är den triviala karaktären modulo $q > 1$, så följer det från beviset av Proposition 21 att

$$L(s, \chi_0) = \zeta(s) \cdot \prod_{p|q} \left(1 - \frac{1}{p^s}\right). \tag{116}$$

Sats 29 *(i) Om χ är en trivial karaktär, då är $L(s, \chi)$ analytisk i $\sigma > 1$ och har en meromorfisk utvidning till $\sigma > 0$ med en enda enkel pol i $s = 1$.*

(ii) Om χ är icke-trivial, då är $L(s, \chi)$ analytisk i $\sigma > 0$.

(iii)

$$L(s, \chi) = \prod_{p} \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}, \tag{117}$$

när båda sidorna konvergerar enligt (i) eller (ii) resp.

Bevis : Part (i) is a consequence of (116) and Theorem 28. Part (iii) follows from Proposition 22. Part (ii) follows from Proposition 12(ii) and eq. (111) with $a_n = \chi(n)$ and $b_n = n^{-s}$, which together imply that the series converges absolutely when $\sigma > 0$ and uniformly in $\sigma \geq \delta > 0$.

Dirichlet's adaption of Euler's method now leads to the following :

Theorem 30 *Let $a, q > 0$ and $GCD(a, q) = 1$. If*

$$L(1, \chi) \neq 0,$$

for every non-trivial character χ modulo q, then there are infinitely many primes $p \equiv a$ (mod q). Moreover, the sum of their reciprocals diverges.

Proof : Let χ be a character modulo q. We consider the L-series

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

[10] Jag tror att denna L betyder 'Legendre', eftersom i de första L-serierna som studerades av Dirichlet var karaktären en Legendre symbol.

By Sats 29, these functions are all meromorphic in $\sigma > 0$, and when $\sigma > 1$ each has an Euler product

$$L(s,\chi) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}.$$

Taking log of both sides and expanding the HL in a Taylor series we obtain, for $\sigma > 1$,

$$\log L(s,\chi) = \sum_p \sum_{m=1}^{\infty} \frac{\chi(p^m)}{mp^{ms}}. \qquad (118)$$

The difference between this and the previous situation is that now we are only interested in those primes $p \equiv a \pmod q$, for some a with $(a,q) = 1$. So how do we isolate these primes in the sum (118) ? The trick is to use the following lemma, which is an immediate consequence of Proposition 12(i) :

Lemma 6 *Let $a, n, q > 0$ with $(a,q) = (n,q) = 1$. Then*

$$\sum_\chi \bar{\chi}(a)\chi(n) = \begin{cases} \phi(q), & \text{if } n \equiv a \ (\text{mod } q), \\ 0, & \text{otherwise}, \end{cases} \qquad (119)$$

where the sum is taken over all characters modulo q.

From (118) and (119) we obtain, by a simple calculation,

$$\frac{1}{\phi(q)} \sum_\chi \bar{\chi}(a) \log L(s,\chi) = \sum_{m=1}^{\infty} \sum_{p^m \equiv a \pmod q} \frac{1}{mp^{ms}}. \qquad (120)$$

Here we're still assuming that $\text{Re}(s) > 1$, and the sum is taken over all characters modulo q. Next, as in the proof of Theorem 27, we split the terms of the sum into two groups, those with $m = 1$ and those with $m > 1$. We observe that the latter sum is bounded as $s \to 1$ and conclude that

$$\lim_{s\to 1^+} \frac{1}{\phi(q)} \sum_\chi \bar{\chi}(a) \log L(s,\chi) = \lim_{s\to 1^+} \sum_{p \equiv a \pmod q} p^{-s} + O(1). \qquad (121)$$

Dirichlet's theorem is (as in the case of Theorem 27) precisely the statement that the limit of the right-hand sum is $+\infty$. Hence we have reduced the proof of the theorem to showing that

$$\lim_{s\to 1^+} \frac{1}{\phi(q)} \sum_\chi \bar{\chi}(a) \log L(s,\chi) = +\infty. \qquad (122)$$

If $\chi = \chi_0$, the trivial character, then it follows from (116) that $L(s, \chi_0) \to +\infty$ as $s \to 1^+$. Hence, (122) would be proven if we could show that $\log L(s, \chi)$ were bounded, as $s \to 1^+$, for every $\chi \neq \chi_0$.

But we know from Sats 29 that if $\chi \neq \chi_0$, then $L(s, \chi)$ is analytic in the range $\mathrm{Re}(s) > 0$. In particular, $L(s, \chi)$ is bounded as $s \to 1$. Hence, by choosing a suitable branch of the logarithm, the same is true of $\log L(s, \chi)$ unless $L(1, \chi) = 0$. This completes the proof of Theorem 30.

Lecture 20 : 171204

We now turn to a proof of Sats 21. This will lead us into the field of *Diophantine approximation theory*, which is concerned (in a sense which we can make precise below) with how 'well' an irrational number can be approximated by rationals whose denominators are 'not too large'. This will in turn lead us to talk a bit about *transcendental numbers*.

The simplest result in Diophantine approximation is

Dirichlets approximationsats *Låt θ_1, θ_2 vara reella tal med $\theta_2 > 1$. Då finns det heltal p, q, med $0 < q < \theta_2$ så att*

$$|q\theta_1 - p| \leq 1/\theta_2. \tag{123}$$

BEVIS : (Se också Baker s.43). Antag först att $\theta_2 \in \mathbf{N}$. För värje q så att $0 < q \leq \theta_2$ sätt

$$r_q := q\theta_1 - [q\theta_1].$$

Detta ger θ_2 reella tal (kanske med repititoner) i intervallen $[0, 1)$. Dela upp denna intervall i θ_2 delintervaller I_t där

$$I_t = \left[\frac{t}{\theta_2}, \frac{t+1}{\theta_2}\right), \qquad t = 0, 1, ..., \theta_2 - 1.$$

Då har vi följande två möjligheter :

Fall I : Det finns $q_1 \neq q_2$ så att både $r_{q_1}, r_{q_2} \in I_t$ för något t. WLOG, $q_1 > q_2$. Då har vi att

$$|q\theta_1 - p| < 1/\theta_2,$$

där $q = q_1 - q_2$, så att $0 < q < \theta_2$, och $p = [q_1\theta_1] - [q_2\theta_2]$.

Fall II : Det finns precis en r_q i varje I_t. Då finns det $q_1 \neq q_2$ så att $r_{q_1} \in I_0$ och $r_{q_2} \in I_{\theta_2 - 1}$. Låt $q = \min\{q_1, q_2\}$. Då är $0 < q < \theta_2$ och $|q\theta_1 - p| \leq 1/\theta_2$ där $p = [q\theta_1]$.

Detta avslutar beviset när $\theta_2 \in \mathbf{N}$. För $\theta_2 \notin \mathbf{N}$, tillämpa resultatet för paret $(\theta_1, [\theta_2]+1)$.

OBS! Det följer från ovanstående bevis att satsen stämmer med en STRÄNG olikhet i (123) om inte $\theta_1 \in \mathbf{Q}$.

Anmärkning En ekvivalent formulering av satsen som gör kopplingen till approximationer av irrationella med rationella tal lite tydligare är

Låt θ vara ett irrationellt tal. Då finns det oändligt många par p, q av relativt prima heltal så att

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^2}.$$

En direkt, och imponerande tillämpning av Dirichlets sats är

Sats 31 *Låt $d > 0$ vara ett heltal som inte är en perfekt kvadrat. Då finns det oändligt många $(x, y) \in \mathbf{Z}^2$ så att $x^2 - dy^2 = 1$.*

BEVIS : (se Baker s.64-65). Låt $m > 0$. Tillämpa Dirichlets approximationsats till paret $\theta_1 = \sqrt{d}$, $\theta_2 = m\sqrt{d}$. Då finns det heltal p_m, q_m, med $0 < q_m < m\sqrt{d}$ så att

$$\left|p_m - q_m\sqrt{d}\right| < \frac{1}{m\sqrt{d}}. \tag{124}$$

Sätt $\alpha_m := p_m - q_m\sqrt{d}$, så att $|\alpha_m| < 1/m\sqrt{d}$. Vi har att

$$\begin{aligned} \alpha_m' = p_m + q_m\sqrt{d} = \alpha_m + 2q_m\sqrt{d} \Rightarrow |\alpha_m'| &\leq |\alpha_m| + 2q_m\sqrt{d} \\ &< \tfrac{1}{m\sqrt{d}} + 2q_m\sqrt{d} \\ &< 3q_m\sqrt{d}, \text{ säg.} \end{aligned}$$

Då har vi att

$$|n(\alpha_m)| = |\alpha_m\alpha_m'| = |\alpha_m|\cdot|\alpha_m'| < \frac{1}{m\sqrt{d}}\cdot 3q_m\sqrt{d} < 3\frac{q_m}{m} < 3\sqrt{d}.$$

Poängen är att $|n(\alpha_m)|$ är begränsad oberoende av m.

Näst, eftersom $\sqrt{d} \notin \mathbf{Q}$ då är varje $\alpha_m \neq 0$. Men $1/m\sqrt{d} \to 0$ då $m \to \infty$, så det måste finnas oändligt många olika tal bland de α_m. Eftersom deras normer är begränsade, hittar vi då oändligt många olika tal av formen $p - q\sqrt{d}$, där $p, q \in \mathbf{Z}$ och $q > 0$, vars normer är alla lika, till N säg. Vidare kan vi nu välja en oändlig delmängd av dessa tal så att alla p är kongruenta till varandra modulo N, och detsamma för alla q. Välj nu två bland dessa tal, säg $\alpha_i = p_i - q_i\sqrt{d}$, för $i = 1, 2$ och betrakta

$$\eta := \frac{\alpha_1}{\alpha_2}.$$

PÅSTÅENDE : $\eta \neq \pm 1$ och $\eta = x + y\sqrt{d}$, där $x, y \in \mathbf{Z}$ och $x^2 - dy^2 = 1$.

Först, $\alpha_1 \neq \alpha_2 \Rightarrow \eta \neq 1$. Att $\eta \neq -1$ följer från att både q_1 och q_2 är positiva. A priori är $\eta = x + y\sqrt{d}$ för några $x, y \in \mathbf{Q}$, och $x^2 - dy^2 = \eta\eta' = n(\eta) = n\left(\frac{\alpha_1}{\alpha_2}\right) = \frac{n(\alpha_1)}{n(\alpha_2)} = 1$. Alltså, kvarstår det att bevisa att $x, y \in \mathbf{Z}$. En explicit beräkning ger

$$\eta = \frac{\alpha_1}{\alpha_2} = \frac{\alpha_1\alpha_2'}{n(\alpha_2)} = \frac{(p_1 - q_1\sqrt{d})(p_2 + q_2\sqrt{d})}{N},$$

så att

$$x = \frac{p_1p_2 - q_1q_2d}{N}, \quad y = \frac{p_1q_2 - q_1p_2}{N}. \tag{125}$$

Det följer nu från faktumet att $p_1 \equiv p_2$ och $q_1 \equiv q_2$ modulo N att täljarna i båda uttrycken i (125) är delbara med N, och alltså att $x, y \in \mathbf{Z}$, v.s.s.v.

Detta räcker nu för att bevisa Sats 31, eftersom α_1 och α_2 valdes från en oändlig mängd, så att det finns oändligt många möjligheter för talet η.

VIKTIG ANMÄRKNING : Ovanstående bevis, tillsammans med beviset av Dirichlets approximationssats, är konstruktiva, men ger en ganska krånglig algoritm för att skapa lösningar till $x^2 - dy^2 = 1$. Denna algoritm kan uppfattas på ett mycket mer elegant sätt med hjälp av *kedjebråk*. Se Kapitel 6 i Bakers bok.