Elementary Number Theory

Franz Lemmermeyer

May 10, 2006

Franz Lemmermeyer
franz@fen.bilkent.edu.tr
http://www.fen.bilkent.edu.tr/∼franz/

Contents

Preface

This introduction to number theory covers the basic material up to the quadratic reciprocity law, and stresses applications to cryptography. In the last chapter, we introduce the theory of binary quadratic forms and complex quadratic number fields.

Chapter 1

The Arithmetic of $\mathbb{Z}$

Number theory deals with properties of the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$, or, more generally, of the integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. In particular it studies objects like prime numbers, whose role in arithmetic parallels that of atoms (in the classical sense) in physics.

1.1 Divisibility

We say that an integer $b \in \mathbb{Z}$ divides $a \in \mathbb{Z}$ (and write $b \mid a$) if there exists an integer $q \in \mathbb{Z}$ such that $a = bq$.

Proposition 1.1. *For all integers a, b, c we have*

1. $a \mid 0$;
2. $1 \mid a$ *and* $a \mid a$;
3. $a \mid b$ *if and only if* $(-a) \mid b$;
4. $a \mid b$ *if and only if* $a \mid (-b)$;
5. *if* $a \mid b$ *and* $b \mid c$, *then* $a \mid c$;
6. *if* $a \mid b$ *and* $a \mid c$, *then* $a \mid (b \pm c)$.

Proof. The proofs are immediate. For showing the fifth claim, observe that we have $b = aq$ and $c = br$ for $q, r \in \mathbb{Z}$; but then $c = br = a(qr)$, hence $a \mid c$.

The last claim is proved as follows: we have $b = aq$ and $c = ar$ for $q, r \in R$; then $b \pm c = a(q \pm r)$ implies that $a \mid (b \pm c)$. □

Elements dividing 1 are called units; the units in $\mathbb{Z}$ are -1 and $+1$. First of all, they are units because they divide 1. Now assume that $r \in \mathbb{Z}$ is a unit; then there exists an element $s \in \mathbb{Z}$ with $rs = 1$. Clearly $r, s \neq 0$, hence $|r|, |s| \geq 1$. If $|r| > 1$, then $0 < |s| < 1$, but there are no integers strictly between 0 and 1. Thus $|r| = 1$, that is, $r = \pm 1$.

Monoids

The notion of divisibility makes sense in any monoid. A monoid is a set M on which a multiplication is defined (a map $M \times M \longrightarrow M$) such that

1. multiplication is commutative and associative;
2. M contains a neutral element ($1 \in M$);
3. M is cancellative: if $xy = xz$ for $x, y, z \in M$, then $x = y$.

Examples for monoids are the nonzero natural numbers, the nonzero integers, as well as e.g. the following sets:

1. $M = \{1, 3, 5, 7, \ldots\} = 2\mathbb{N}_0 + 1$;
2. $M = \{1, 2, 4, 6, \ldots\} = 2\mathbb{N} \cup \{1\}$;
3. $M = \{1, 4, 7, 10, \ldots\} = 3\mathbb{N}_0 + 1$; this example is due to Hilbert.
4. $M = \{1, 2, 4, 8, 16, \ldots\}$.

If R is any domain (a ring without zero divisors, that is, with the property that $ab = 0$ in R implies $a = 0$ or $b = 0$), then $M = R$ and $M = R \setminus \{0\}$ are monoids.

The following properties of divisibbility known from $\mathbb{Z}$ hold more generally for monoids:

Proposition 1.2. *For all elements a, b, c of a monoid, we have*

1. $1 \mid a$ *and* $a \mid a$;
2. *if* $a \mid b$ *and* $b \mid c$, *then* $a \mid c$;
3. *if* $a \mid b$ *then* $a \mid bc$.

The proofs are easy.

For our next result, recall that a group is a set G endowed with a composition $G \times G \longrightarrow G$ called multiplication (the composition of $a, b \in G$ will be denoted by ab), such that the following properties are verified:

1. Existence of a unit: there is an element $1 \in G$ such that $1a = a$ for all $a \in G$.
2. Existence of an inverse: for every $a \in G$ there is a $b \in G$ such that $ab = 1$.
3. Associativity: we have $a(bc) = (ab)c$ for all $a, b, c \in G$.

Proposition 1.3. *The set $M^\times$ of units (elements dividing 1) in some monoid M is a group.*

Proof. We have to check the axioms. First, $1 \in M^\times$ shows the existence of a neutral element. If u is a unit, then by definition there is some $v \in M$ such that $uv = 1$. But then $v = u^{-1}$ is also a unit, so inverses exist. Finally, if u and v are units, then $uu' = vv' = 1$ for some $u', v' \in M$, and then $(uv)(u'v') = 1$, hence uv is a unit.

Note that commutativity and associativity are inherited from M: if these axioms hold for all elements in M, then they surely will hold for all elements in $M^\times$. □

We now give two important definitions. Let M be a monoid; then a non-unit $p \in M$ is called

- irreducible if it only has trivial factorizations, i.e. if $p = ab$ for $a, b \in M$ implies that $a \in M^\times$ or $b \in M^\times$.
- prime if $p \mid ab$ for $a, b \in M$ implies that $p \mid a$ or $p \mid b$.

The following observation will be used repeatedly:

Lemma 1.4. *If p and q are irreducible elements with $p \mid q$, then $q = pu$ for some unit u.*

Proof. Assume that $p \mid q$ and write $q = pu$. Since q is irreducible, it only has trivial factorizations, hence p or u must be a unit. But p is irreducible, hence a nonunit by definition. Thus u is a unit. □

Being prime is a stronger property than being irreducible:

Proposition 1.5. *In monoids, primes are irreducible.*

Proof. Let p be prime. We want to show it's irreducible, so assume that $p = ab$; we have to prove that a or b is a unit. Now clearly $p \mid ab$, and since p is prime, we have $p \mid a$ or $p \mid b$. Assume without loss of generality that $p \mid a$. Then $a = pc$ for some $c \in M$, hence $p = ab = pbc$, and since M is cancellative we deduce that $1 = bc$. Thus b is a unit, and this concludes the proof. □

It is not true at all that irreducibles are always prime. It is basically in order to have lots of examples that we have dealt with monoids here. Consider e.g. the monoid $M = \{1, 2, 4, 6, \ldots\}$; here 2 is irreducible since clearly $2 = 1 \cdot 2 = 2 \cdot 1$ are the only factorizations of 2. On the other hand, 2 is not prime: we have $2 \mid 6 \cdot 6$ since $36 = 2 \cdot 18$, but $2 \nmid 6$ because 6 is not divisible by 2 in M. In fact, M does not have any primes at all: 1 is a unit, elements of the form $4n = 2 \cdot 2n$ for $n \in \mathbb{N}$ are reducible, and elements of the form $4n + 2$ for $n \in \mathbb{N}$ are irreducible (products of elements $2n$ and $2m$ have the form $4mn$) but not prime: we have $2(2n + 1) \mid 6(2n + 1) \cdot 6(2n+^)$, but $2(2n + 1) \nmid 6(2n + 1)$ because $2 \nmid 6$.

1.2 Unique Factorization

In this section we will show that every nonzero integer can be written (in an esentially unique way) as a product of irreducibles. This is not obvious, as the monoid $M = \{1, 5, 9, \ldots\}$ of natural numbers of the form $4n + 1$ shows: here $21 \cdot 21 = 9 \cdot 49$ are two distinct factorizations of 441 into irreducibles. In abstract algebra you will learn that nonunique factorization is connected to the existence of irreducibles that are not prime.

Theorem 1.6. *Every nonzero integer $n \in \mathbb{Z}$ can be written as the product of a unit and positive irreducible elements. This product is unique up to the order of the factors.*

The following proof goes back to Zermelo:

Proof. We first remark that it is sufficient to prove everything for natural numbers, since every nonzero integer is ± 1 times a natural number.

Existence: This is trivial for $m = 1, 2, 3$. Assume we have proved the existence of a factorization into irreducibles for all numbers $< m$. If n is irreducible, there is nothing to prove; if not, then $n = ab$ for natural numbers $a, b < m$. These can be factored into irreducibles by induction assumption, hence the same is true for their product.

Uniqueness: Clearly, 2, 3, and $4 = 2 \cdot 2$ have unique factorizations into irreducibles. Assume we have proved uniqueness for all numbers $< m$; we then show by induction that n also has unique factorization.

In fact, let p be the smallest irreducible factor of m; then $m = ph$, and $h < m$ implies that h has unique factorization (into irreducibles). Now assume there is a second factorization of m, and let q denote the smallest irreducible factor occurring there (in the monoid $M = \{1, 4, 7, 10, \ldots\}$, the example $m = 100 = 4 \cdot 25 = 10 \cdot 10$ would give $p = 4$ and $q = 10$). Write $m = qk$. Since $p \leq q$ by the minimality of p, there are two cases:

1. $p = q$. Then $h = k < m$, hence h and k can be factored uniquely, and the factorizations $m = pk = qh$ are not different, contradicting our assumption.

2. $p < q$ (this happens in the example $4 \cdot 25 = 10 \cdot 10$). Then

$$n := m - pk = \begin{cases} qk - pk & = (q - p)k \\ ph - pk & = p(h - k) \end{cases}.$$

Since n is positive and smaller than m, it has a unique factorization into irreducibles. The second factorization shows that p occurs among its factors, hence p must occur among the irreducible factors of $q - p$ or k. But the irreducible factors of qk are all $\geq q$ by the choice of q, hence all the irreducible factors of k are $\geq q > p$. This means that p must occur as a factor of $q - p$, i.e., $p \mid (q - p)$. But then p must divide $q = (q - p) + p$,

and this implies $p = q$ by Lemma 1.4. This contradiction proves unique factorization.

□

Corollary 1.7. *In $\mathbb{Z}$, irreducibles are prime.*

In the traditional proofs of unique factorization, this result is proved first and then used to prove unique factorization.

Proof. Let $p \in \mathbb{N}$ be irreducible, and assume that $p \mid ab$. Write $a = up_1 \cdots p_r$ and $b = vq_1 \cdots q_s$ for units $u, v \in \{-1, +1\}$ and irreducibles p_i and q_j. Then $ab = uvp_1 \cdots p_r q_1 \cdots q_s$ is the factorization of ab into irreducibles. On the other hand, $ab = pc$, and by factoring c into irreducibles we see that p occurs in the factorization of ab into irreducibles. Thus p must be one of the p_i or one of the q_j. But then $p \mid a$ in the first, and $p \mid b$ in the second case. □

In the monoid $\{1, 2, 4, 8, \ldots\}$ consisting of powers of 2, there is only one irreducible element, namely 2. In the monoid $M = \{1, 2, 4, 6, \ldots\}$, proving the existence of infinitely many irreducibles was easy, since we could write them down explicitly. In $\mathbb{Z}$, the corresponding property is more difficult to prove:

Proposition 1.8. *In $\mathbb{Z}$, there exist infinitely many primes.*

The following famous proof is due to Euclid (ca 300 BC):

Proof. Examples of primes are 2, 3, 5. Assume that there are only finitely many primes $p_1 \cdots p_n$. Consider the integer $N = p_1 \cdots p_n + 1$. Then $N \neq 0$ is not a unit in $\mathbb{Z}$ because $N > 1$. But then N must be divisible by some prime p. If $p = p_j$, then $p_j \mid N$ and $p_j \mid N - 1$ (since $N - 1 = p_1 \cdots p_n$), and thus $p_j \mid 1$ (the difference between N and $N - 1$). But then p_j is a unit, and this is a contradiction. □

Note that Euclid's proof does not show that the numbers $N = p_1 \cdots p_n + 1$ are prime. With `pari`, it is easy to factor the first few such numbers: typing in `factor(2*3*5*7+1)`, for example, will yield the answer 211. In this way, the following table was constructed:

$$\begin{aligned}
2 + 1 &= 3 \\
2 \cdot 3 + 1 &= 7 \\
2 \cdot 3 \cdot 5 + 1 &= 31 \\
2 \cdot 3 \cdot 5 \cdot 7 + 1 &= 211 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 &= 2311 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 &= 59 \cdot 509 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 + 1 &= 19 \cdot 97 \cdot 277
\end{aligned}$$

1.3 Congruences

Congruences are a very clever notation invented by Gauss (and published in 1801 in his "Disquisitiones Arithmeticae") to denote the residue of a number a upon division by a nonzero integer m. More precisely, he wrote $a \equiv b \bmod m$ if $m \mid (a-b)$. for elements $a, b, m \in \mathbb{Z}$.

Examples.

- $10 \equiv 3 \bmod 7$;
- $10 \equiv 0 \bmod 5$;
- $5 \equiv 2 \equiv -1 \bmod 3$.

The rules for divisibility can now be transferred painlessly to congruences: first we observe

Proposition 1.9. *Congruence between integers is an equivalence relation.*

Proof. Recall that a relation is called an equivalence relation if it is reflexive, symmetric and transitive. In our case, we have to show that the relation $\equiv$ has the following properties:

- reflexivity: $a \equiv a \bmod m$;
- symmetry: $a \equiv b \bmod m$ implies $b \equiv a \bmod m$;
- transitivity: $a \equiv b \bmod m$ and $b \equiv c \bmod m$ imply $a \equiv c \bmod m$

for $a, b, c \in \mathbb{Z}$ and $m \in \mathbb{Z} \setminus \{0\}$.

The proofs are straightforward. In fact, $a \equiv a \bmod m$ means $m \mid (a-a)$, and every integer $m \neq 0$ divides 0. Similarly, $a \equiv b \bmod m$ is equivalent to $m \mid (a-b)$; but this implies $m \mid (b-a)$, hence $b \equiv a \bmod m$. Finally, if $a \equiv b \bmod m$ and $b \equiv c \bmod m$, then $m \mid (b-a)$ and $m \mid (c-b)$, hence m divides the sum $c-a = (c-b)+(b-a)$, and we find $a \equiv c \bmod m$ as claimed. □

Since $\equiv$ defines an equivalence relation, it makes sense to talk about equivalence classes. The equivalence class $[a]$ (or $[a]_m$ if we want to express the dependence on the modulus m) of an integer a consists of all integers $b \in \mathbb{Z}$ such that $b \equiv a \bmod m$; in particular, every residue class contains infinitely many integers. In the special case $m = 3$, for example, we have

$$\begin{aligned} [0] &= \{\ldots, -6, -3, 0, 3, 6, \ldots\}, \\ [1] &= \{\ldots, -5, -2, 1, 4, 7, \ldots\}, \\ [2] &= \{\ldots, -4, -1, 2, 5, 8, \ldots\}, \\ [3] &= \{\ldots, -3, 0, 3, 6, 9, \ldots\} = [0], \end{aligned}$$

etc. Note that $[0] = [3] = [6] = \ldots$ (in fact, $[0] = [a]$ for any $a \in [0]$), and similarly $[1] = [4] = \ldots$. In general, we have $[a] = [a']$ if and only if $a \equiv a' \bmod m$, that is, if and only if $m \mid (a - a')$.

In the case $m = 3$, there were exactly 3 different residue classes modulo 3, namely [0], [1], and [2] (or, say, [0], [1], and $[-1]$ since $[-1] = [2]$). This holds in general:

Lemma 1.10. *For any integer $m > 1$, there are exactly m different residue classes modulo m, namely $[0]$, $[1]$, $[2]$, $\ldots$, $[m-1]$.*

Proof. We first show that these classes are pairwise distinct. To this end, assume that $[a] = [b]$ for $0 \leq a, b < m$; this implies $b \in [a]$, hence $a \equiv b \bmod m$ or $m \mid (b - a)$: but since $|b - a| < m$, this can only happen if $a = b$.

Next, there are no other residue classes: given any class $[a]$, we write $a = mq + r$ with $0 \leq r < m$ (the division algorithm at work again), and then $[a] = [r]$ is one of the classes listed above. □

The set $\{0, 1, 2, \ldots, m-1\}$ is often called a complete set of representatives modulo m for this reason. Sometimes we write $r + m\mathbb{Z}$ instead of $[r]$.

The one thing that makes congruences *really* useful is the fact that we can define a ring structure on the set of residue classes. This is fundamental, so let us do this in detail.

The elements of our ring $\mathbb{Z}/m\mathbb{Z}$ will be the residue classes $[0]$, $[1]$, $\ldots$, $[m-1]$ modulo m. We have to define an addition and a multiplication and then verify the ring axioms.

• Addition $\oplus$: Given two classes $[a]$ and $[b]$, we put $[a] \oplus [b] = [a + b]$. We have to check that this is well defined: assume that $[a] = [a']$ and $[b] = [b']$; then we have to show that $[a + b] = [a' + b']$. But this is easy: we have $a - a' \in m\mathbb{Z}$, say $a - a' = mA$, and similarly $b - b' = mB$. But then $(a + b) - (a' + b') = m(A + B) \in m\mathbb{Z}$, hence $[a + b] = [a' + b']$.

The neutral element is the residue class $[0] = m\mathbb{Z}$, and the inverse element of $[a]$ is $[-a]$, or, if you prefer, $[m - a]$. In fact, we have $[a] \oplus [0] = [a + 0] = [a]$ and $[a] \oplus [-a] = [a + (-a)] = [0]$. The law of associativity and the commutativity are inherited from the corresponding properties of integers: since e.g. $(a + b) + c = a + (b + c)$, we have $([a] \oplus [b]) \oplus [c] = [a] \oplus ([b] \oplus [c])$.

• Multiplication $\odot$: of course we put $[a] \odot [b] = [ab]$. The verification that this is well defined is left as an exercise. The neutral element is the class $[1]$.

• Distributive Law: Again, $([a] \oplus [b]) \odot [c] = [a] \odot [c] \oplus [b] \odot [c]$ follows from the corresponding properties of integers.

Theorem 1.11. *The residue classes $[0]$, $[1]$, $\ldots$, $[m-1]$ modulo m form a ring $\mathbb{Z}/m\mathbb{Z}$ with respect to addition $\oplus$ and multiplication $\odot$.*

Now that we have introduced the rings that we will study for some time to come, we simplify the notation by writing $+$ and $\cdot$ instead of $\oplus$ and $\odot$. Moreover, we will drop our references to classes and deal only with the integers representing them; in order to make clear that we are dealing with residue classes, we write $\equiv$ instead of $=$ and add a "mod m" at the end. What this means in practice is that we identify $\mathbb{Z}/m\mathbb{Z}$ with the set of integers $\{0, 1, \ldots, m-1\}$.

Applications: ISBN (International Standard Book Number)

From the 1970s onward books are assigned an ISBN consisting of four parts: the first block specifies the country (or rather the language of the country), the second block gives information about the publishing company, the third about the book within that company, and the last digit is a check digit that is computed as follows: multiply the digits of the ISBN by 1, 2, 3, ..., 10, starting on the left; the check digit is the integer ≤ 10 for which the sum of these products is $\equiv 0 \bmod 11$. The check 'digit' X stands for 10.

Example: compute the check digit of the ISBN 0-387-94225-?. We find $1\cdot 0+2\cdot 3+3\cdot 8+4\cdot 7+5\cdot 9+6\cdot 4+7\cdot 2+8\cdot 2+9\cdot 5+10? \equiv 4+10? \bmod 11$, and since $10 \equiv -1 \bmod 11$, this gives $4-? \equiv 0 \bmod 11$, so $? = 4$, and the complete ISBN is 0-387-94225-4.

It is easy to see that if you type in an ISBN and make a single error, then the check digit will catch it; thus the ISBN is an example of a 1-error detecting code.

1.4 Greatest Common Divisors in $\mathbb{Z}$

We will now introduce greatest common divisors: we say that d is a greatest common divisor of $a, b \in \mathbb{Z}$ and write $d = \gcd(a, b)$ if d satisfies the following two properties:

1. $d \mid a$, $d \mid b$: d is a common divisor of a and b.
2. if $e \in \mathbb{Z}$ satisfies $e \mid a$ and $e \mid b$, then $e \mid d$: every common divisor of a and b divides d.

We can use the unique factorization property to give a formula for the gcd of two integers. Before we do so, let us introduce some notation. We can write an $a \in \mathbb{Z}$ as a product of primes. In fact fact we can write $a = \pm \prod p_i^{a_i}$, where the product is over all irreducible elements p_1, p_2, p_3, ..., and where at most finitely many a_i are nonzero. In order to avoid the $\pm$ in our formulas, let us restrict to positive integers from now on.

Lemma 1.12. *For integers $a, b \in \mathbb{N}$ we have $b \mid a$ if and only if $b_i \leq a_i$ for all i, where $a = \prod p_i^{a_i}$ and $b = \prod p_i^{b_i}$ are the prime factorizations of a and b.*

Proof. We have $b \mid a$ if and only if there is a $c \in \mathbb{N}$ such that $a = bc$. Let $c = \prod p_i^{c_i}$ be its prime factorization. Then $c_i \geq 0$ for all i, and $a_i = b_i + c_i$, hence $b \mid a$ is equivalent to $a_i \geq b_i$ for all i. □

Here's our formula for gcd's:

Theorem 1.13. *The gcd of two nonzero integers*

$$a = \prod p_1^{a_i} \quad \text{and } b = \prod p_i^{b_i}$$

is given by

$$d = \prod p_i^{\min\{a_i, b_i\}}.$$

Proof. We have to prove the two properties characterizing gcd's:

1. $d \mid a$ and $d \mid b$. But this follows immediately from Lemma 1.12.

2. If $d' \mid a$ and $d' \mid b$, then $d' \mid d$. In fact, write down the prime factorization $d' = \prod p_i^{d'_i}$ of d'. Then $d' \mid a$ and $d' \mid b$ imply $d'_i \leq \min(a_i, b_i) = d_i$, hence $d' \mid d$.

Now assume that d and d' are gcd's of a and b. Then $d \mid d'$ by 2. since d' is a gcd, and $d' \mid d$ since d is a gcd, hence $d' = \pm d$. □

For the ring $\mathbb{Z}$ of integers, we have much more than the mere existence of gcd's: the gcd of two integers $a, b \in \mathbb{Z}$ has a "Bezout representation",[1] that is, if $d = \gcd(a, b)$, then there exist integers $m, n \in \mathbb{Z}$ such that $d = am + bn$.

Theorem 1.14 (Bezout's Lemma)**.** *Assume that $d = \gcd(a, b)$ for $a, b \in \mathbb{Z}$; then d has a Bezout representation.*

Proof. Consider the set $D = a\mathbb{Z} + b\mathbb{Z} = \{am + bn : m, n \in \mathbb{Z}\}$. Clearly D is a nonempty set, and if $c \in D$ then we also have $-c \in D$. In particular, D contains positive integers.

Let d be the smallest positive integer in D; we claim that $d = \gcd(a, b)$. There are two things to show:
Claim 1: d is a common divisor of a and b. By symmetry, it is sufficient to show that $d \mid a$. Write $a = rd + s$ with $0 \leq s < d$; from $d = am + bn$ we get $s = a - rd = a - r(am + bn) = a(1 - rm) + b(-rn)$, hence $s \in D$. The minimality of d implies $s = 0$, hence $d \mid m$.
Claim 2: if e is a common divisor of a and b, then $e \mid d$. Assume that $e \mid a$ and $e \mid b$. Since $d = am + bn$, we conclude that $e \mid d$.

The existence of the Bezout representation is a simple consequence of the fact that $d \in D$. □

Note that the key of the proof is the existence of a division with remainder.

Bezout's Lemma can be used to give an important generalization of the property $p \mid ab \Longrightarrow p \mid a$ or $p \mid b$ of primes p:

Proposition 1.15. *If $m \mid ab$ and $\gcd(m, b) = 1$, then $m \mid a$.*

Proof. Write $ab = mn$; by Bezout, there are $x, y \in \mathbb{Z}$ such that $mx + by = 1$. Multiplying through by a gives $a = max + aby = max + mny = m(ax + ny)$, that is, $m \mid a$. □

Note that our proof of Bezout's result did not use unique factorization. In fact, we can use Bezout to give another proof of unique factorization. First we show that in $\mathbb{Z}$, irreducibles are prime.

[1] Etienne Bezout: 1730 (Nemours, France) – 1783 (Basses-Loges, France)

Second Proof of Cor. 1.7. In fact, let $p \in \mathbb{N}$ be irreducible, and assume that $p \mid ab$. Let $d = \gcd(p, b)$; since $d \mid p$ and p is irreducible, we either have $d = 1$ or $d = p$. If $d = p$, then $p \mid b$. If $d = 1$, then Bezout gives us $x, y \in \mathbb{Z}$ with $1 = px + by$. Multiplying through by a shows $a = pax + aby$. Since p divides both terms on the right hand side, we conclude that $p \mid a$. □

Now we can give

Second Proof of Unique Factorization. Assume it fails, and let n be the smallest positive integer with two distinct factorizations (the smallest criminal, as such elements are called in group theoretic circles) into positive irreducible elements. Write $n = p_1 \cdots p_r = q_1 \cdots q_s$. Then $p_1 \mid q_1 \cdots q_s$, and since irreducibles are prime, p_1 must divide a factor of the right hand side, say $p_1 \mid q_1$. By Lemma 1.4 we see that $q_1 = p_1 u$ for some unit u, and since $p_1, q_1 > 0$ by assumption we must have $u = 1$ and $p_1 = q_1$. But then cancelling p_1 gives $p_2 \cdots p_r = q_1 \cdots q_s$. Since n was the smallest integer with two factorizations, we must have $r = s$ and, possibly after reordering the factors, $p_2 = q_2, \ldots, p_r = q_r$. Thus the two factorizations were the same after all. □

There are a couple of simple properties of the gcd that can easily be proved from the definition or using Bezout or unique factorization:

Proposition 1.16. *For integers $a, m, n \in \mathbb{N}$ we have*

1. $\gcd(am, an) = a \gcd(m, n)$;
2. *If* $a = \gcd(m, n)$, *then* $\gcd(\frac{m}{a}, \frac{m}{b}) = 1$;
3. *if* $a \mid m$ *and* $a \mid n$, *then* $a \mid \gcd(m, n)$;
4. $\gcd(a, mn) \mid \gcd(a, m) \cdot \gcd(a, n)$;
5. *if* $\gcd(a, m) = 1$, *then* $\gcd(a, mn) = \gcd(a, n)$.

Finally, observe that canceling factors in congruences is dangerous: we have $2 \equiv 8 \bmod 6$, but not $1 \equiv 4 \bmod 6$. Here's what we're allowed to do:

Proposition 1.17. *If* $ac \equiv bc \bmod m$, *then* $a \equiv b \bmod \frac{m}{\gcd(m,c)}$.

Proof. We have $m \mid (ac - bc) = c(a - b)$. Write $d = \gcd(m, c)$, $m = dm'$, $c = dc'$, and note that $\gcd(m', c') = 1$. From $dm' \mid dc'(a - b)$ we deduce immediately that $m' \mid c'(a - b)$; since $\gcd(m', c') = 1$, we even have $m' \mid (a - b)$ by Prop 1.15, i.e. $a \equiv b \bmod \frac{m}{\gcd(m,c)}$. □

1.5 The Euclidean Algorithm

In most modern textbooks, Unique Factorization is proved using the Euclidean algorithm; it has the advantage that a similar proof can also be used for other rings, e.g. polynomial rings $K[X]$ over fields K. The Euclidean algorithm is a procedure that computes the gcd of integers without using their prime factorization (which may be difficult to obtain if the numbers involved are large). Moreover, it allows us to compute a Bezout representation of this gcd (note that our proof of Thm. 1.14 was an existence proof, giving no hint at how to compute such a representation).

Given integers m and n, there are uniquely determined integers q_1 and r_1 such that $m = q_1 n + r_1$ and $0 \leq r_1 < n$. Repeating this process with n and r_1, we get $n = r_1 q_2 + r_2$ with $0 \leq r_2 < r_1$, etc. Since $n > r_1 > r_2 > \ldots \geq 0$, one of the r_i, say r_{n+1}, must eventually be 0:

$$\begin{aligned} m &= q_1 n + r_1 && (1.1)\\ n &= q_2 r_1 + r_2 && (1.2)\\ r_1 &= q_3 r_2 + r_3 && (1.3)\\ &\ldots \\ r_{n-2} &= q_n r_{n-1} + r_n && (1.4)\\ r_{n-1} &= q_{n+1} r_n && (1.5) \end{aligned}$$

Example: $m = 56$, $n = 35$

$$\begin{aligned} 56 &= 1 \cdot 35 + 21\\ 35 &= 1 \cdot 21 + 14\\ 21 &= 1 \cdot 14 + 7\\ 14 &= 2 \cdot 7 \end{aligned}$$

Note that the last r_i that does not vanish (namely $r_3 = 7$) is the gcd of m and n. This is no accident: we claim that $r_n = \gcd(m, n)$ in general. For a proof, we have to verify two things:

Claim 1: r_n is a common divisor of m and n. Equation (1.5) shows $r_n \mid r_{n-1}$; plugging this into (1.4) we find $r_n \mid r_{n-2}$, and going back we eventually find $r_n \mid r_1$ from (1.3), $r_n \mid n$ from (1.2) and finally $r_n \mid m$ from (1.1). In particular, r_n is a common divisor of m and n.

Claim 2: if e is a common divisor of m and n, then $e \mid r_n$. This is proved by reversing the argument above: (1.1) shows that $e \mid r_1$, (1.2) then gives $e \mid r_2$, and finally we find $e \mid r_n$ from (1.5) as claimed.

The Euclidean algorithm does more than just compute the gcd: take our example $m = 56$ and $n = 35$; writing the third line as $\gcd(m, n) = 7 = 21 - 1 \cdot 14$ and replacing the 14 by $14 = 35 - 1 \cdot 21$ coming from the second line we get $\gcd(m, n) = 21 - 1 \cdot (35 - 1 \cdot 21) = 2 \cdot 21 - 1 \cdot 35$. Now $21 = 56 - 1 \cdot 35$ gives $\gcd(m, n) = 2 \cdot (56 - 1 \cdot 35) - 1 \cdot 35 = 2 \cdot 56 - 3 \cdot 35$, and we have found a Bezout representation of the gcd of 56 and 35.

This works in complete generality: (1.4) says $r_n = r_{n-2} - q_n r_{n-1}$; the line before, which $r_{n-1} = r_{n-3} - q_{n-1} r_{n-2}$, allows us to express r_n as a $\mathbb{Z}$-linear combination of r_{n-2} and r_{n-3}, and going back we eventually find an expression of r_n as a $\mathbb{Z}$-linear combination of a and b.

Bezout representations have an important practical application: they allow us to compute multiplicative inverses in $\mathbb{Z}/p\mathbb{Z}$. In fact, let $[a]$ denote a nonzero residue class modulo p; since $\mathbb{Z}/p\mathbb{Z}$ is a field, $[a]$ must have a multiplicative inverse, that is, there must be a residue class $[b]$ such that $[ab] = [1]$. Since there are only finitely many residue classes, this can always be done by trial and error (unless p is large): for example, let us find the multiplicative inverse of $[2]$ in $\mathbb{Z}/5\mathbb{Z}$: multiplying $[2]$ successively by $[1]$, $[2]$, $[3]$, $[4]$ we find $[2] \cdot [3] = [6] = [1]$; thus $[2]^{-1} = [3]$ (we occasionally also write $\frac{1}{2} \equiv 3 \bmod 5$).

Computing the inverse of $[2]$ in $\mathbb{Z}/p\mathbb{Z}$ is actually always easy: note that we want an integer b such that $[2b] = [1]$; but $[1] = [p+1]$, hence we can always take $b = \frac{p+1}{2}$.

In general, however, computing inverses is done using Bezout representations. Assume that $\gcd(a, p) = 1$ (otherwise there is no multiplicative inverse), compute integers $x, y \in \mathbb{Z}$ such that $1 = ax + py$; reducing this equation modulo p gives $1 \equiv ax \bmod p$, i.e., $[a][x] = [1]$, or $[a]^{-1} = [x]$.

1.6 Pythagorean Triples

As an application of unique factorization we will now solve several diophantine equations. We will start with one of the oldest diophantine equations, namely

$$x^2 + y^2 = z^2.$$

Integral solutions of this equation are called Pythagorean triples, the most famous being $(x, y, z) = (3, 4, 5)$. It is very easy to write down solutions: the solutions

$$x = m^2 - n^2, \quad y = 2mn, \quad z = m^2 + n^2$$

were known to Euclid, special cases to the Pythagoreans. It is more difficult (and absolutely necessary for our applications to Fermat's Last Theorem) to show that every solution is given by these formulas.

Assume that (x, y, z) is a Pythagorean triple. If d divides two of these, it divides the third, and then $(x/d, y/d, z/d)$ is another Pythagorean triple. We may therefore assume that x, y and z are pairwise coprime; such triples are called primitive.

Now let (x, y, z) be a primitive Pythagorean triple. Clearly, x and y cannot both be even; we now claim that they cannot both be odd. In fact, if $x \equiv y \equiv 1 \bmod 2$, then $x^2 \equiv y^2 \equiv 1 \bmod 4$, hence $z^2 = x^2 + y^2 \equiv 2 \bmod 4$. But squares are always congruent to 0 or 1 mod 4, so this is a contradiction.

Interchanging x and y if necessary we may assume that y is even. Now we transfer the additive problem $x^2 + y^2 = z^2$ into a multiplicative one (if we are to use unique factorization, we need products, not sums) by writing $y^2 = z^2 - x^2 = (z - x)(z + x)$.

Here we have two factors whose product is a square. Unique factorization allows us to derive information from such a situation:

Proposition 1.18. *Let $a, b \in \mathbb{N}$ be coprime integers such that ab is a square. Then a and b are squares.*

Proof. Write down the prime factorizations of a and b as

$$a = p_1^{a_1} \cdots p_r^{a_r}, \quad b = q_1^{b_1} \cdots q_s^{b_s}.$$

Now a and b are coprime, so the set of p_i and the set of q_j are disjoint, and we conclude that the prime factorization of ab is given by

$$ab = p_1^{a_1} \cdots p_r^{a_r} q_1^{b_1} \cdots q_s^{b_s}.$$

Since ab is a square, all the exponents in the prime factorization of ab must be even. This implies that the a_i and the b_j are even, therefore a and b are squares. □

Unfortunately, $z - x$ and $z + x$ are not coprime: since both x and z are odd, their difference and sum must be even, hence they have a common divisor, namely 2. Thus we have to modify our proposition:

Corollary 1.19. *Let $a, b \in \mathbb{N}$ be integers with $\gcd(a, b) = d$ such that ab is a square. Then a/d and b/d are squares.*

Proof. We have $a = dA$ and $b = dB$ with $\gcd(A, B) = 1$ and $AB = (m/d)^2$. Now the proposition gives $A = r^2$ and $B = s^2$. □

We now claim that $\gcd(z - x, z + y) = 2$. We have already seen that 2 is a common divisor, hence it is sufficient to show that $d = \gcd(z-x, z+y) \mid 2$. But $d \mid z - x$ and $d \mid z + x$ implies $d \mid 2z$ and $d \mid 2x$ (take the sum and difference), hence $d = \gcd(2x, 2z) = 2\gcd(x, z) = 2$.

Applying Cor. 1.19 we find that there exist $m, n \in \mathbb{N}$ such that $z - x = 2n^2$ and $z + x = 2m^2$. Adding and subtracting these equations gives $z = m^2 + n^2$ and $x = m^2 - n^2$, and from $y^2 = (z - x)(z + x) = 4m^2n^2$ and $y \in \mathbb{N}$ we deduce that $y = 2mn$.

Note that we must have $\gcd(m, n) = 1$: in fact, any common divisor of m and n would divide x, y and z contradicting our assumption that our triple be primitive. We have shown:

Theorem 1.20. *If (x, y, z) is a primitive Pythagorean triple with y even, then there exist coprime integers $m, n \in \mathbb{N}$ such that $x = m^2 - n^2$, $y = 2mn$, and $z = m^2 + n^2$.*

Euler's Trick

The same technique we used for solving $x^2 + y^2 = z^2$ can be used to solve equations of the type $x^2 + ay^2 = z^2$: just write the equation in the form $ay^2 = (z - x)(z + x)$ and use unique factorization.

Equations like $x^2 + y^2 = 2z^2$ at first seem intractable using this approach because we can't produce a difference of squares. Euler, however, saw that in this case multiplication by 2 saves the day because $(2z)^2 = 2x^2 + 2y^2 = (x + y)^2 + (x - y)^2$, hence $(2z - x - y)(2z + x + y) = (x - y)^2$, and now the solution proceeds similarly as for Pythagorean triples.

Let us now show that we can do something similar for any equation of type $AX^2 + BY^2 = CZ^2$ having at least one solution. First, multiplying through by A shows that it is sufficient to consider equations $X^2 + aY^2 = bZ^2$. Assume that (x, y, z) is a solution of this equation. Then

$$\begin{aligned}(bzZ)^2 &= bz^2X^2 + abz^2Y^2 \\ &= (x^2 + ay^2)X^2 + (ax^2 + a^2y^2)Y^2 \\ &= (xX + ayY)^2 + a(yX - xY)^2.\end{aligned}$$

Thus $a(yX - xY)^2 = (bzZ)^2 - (xX + ayY)^2$ is a difference of squares, and we can proceed as for Pythagoren triples. We have proved:

Theorem 1.21. *If the equation $ax^2 + by^2 = cz^2$ has a nontrivial solution in integers, then this equation can be factored over the integers (possibly after multiplying through by a suitable integer).*

1.7 Fermat's Last Theorem

Fermat's Last Theorem claims that the only integral solutions of the diophantine equation

$$x^n + y^n = z^n$$

for $n > 2$ are trivial, i.e., satisfy $xyz = 0$. Its name comes from the fact that at the beginning of the 19th centiry, all of Fermat's claims had either been proved or disproved, with the exception of the one above.

It was realized early on that it is sufficient to prove the theorem for $n = 4$ and for odd primes p. In fact, every integer $n > 2$ is divisible by p or by 4, say $n = pm$ or $n = 4m$. But now $x^n + y^n = z^n$ can be written as $(x^m)^p + (y^m)^p = (z^m)^p$ or $(x^m)^4 + (y^m)^4 = (z^m)^4$, and if the latter equations have only trivial solution, then so does the Fermat equation for exponent n.

In the following we will give a prove for exponent $n = 4$; Kummer's proof for odd primes $p < 100$ requires a good amount of algebraic number theory, and Wiles' proof for all primes $p \geq 5$ is impossible to understand without a huge background in algebra, number theory and algebraic geometry.

The solution of $x^2 + y^2 = z^2$ will help us prove that the diophantine equation

$$X^4 + Y^4 = Z^4 \tag{1.6}$$

has only trivial solutions, namely those with $X = 0$ or $Y = 0$. As a matter of fact, it is a lot easier to prove more, namely that

$$X^4 + Y^4 = Z^2 \tag{1.7}$$

has only trivial solutions (this *is* more: if $X^4 + Y^4$ cannot be a square, it cannot be a fourth power). The proof is quite involved and uses a technique that Fermat called infinite descent.

Infinite Descent

In a nutshell, the idea behind infinite descent is the following: if we want to prove that a certain diophantine equation is impossible in $\mathbb{N}$, it is sufficient to show that for every solution in natural numbers there is another solution that is "smaller", which eventually leads to a contradiction because there is no natural number smaller than 1.

Here are some simple examples:

Proposition 1.22. *The diophantine equation $x^2 + y^2 = 3z^2$ does not have solutions in natural numbers.*

Proof. Assume there are natural numbers $x, y, z > 0$ such that $x^2 + y^2 = 3z^2$. Then $x^2 + y^2 \equiv 0 \pmod 3$. The following table gives the congruence class of $x^2 + y^2$ modulo 3 in terms of x and y:

	0	1	2
0	0	1	1
1	1	2	2
2	1	2	2

Note that, since $2^2 \equiv 1^2 \bmod 3$, the last row and column were actually superfluous. What this table is showing is that if $x^2 + y^2 \equiv 0 \bmod 3$, then $x \equiv y \equiv 0 \bmod 3$. A quicker way of seeing this is provided by the following argument: if $3 \nmid y$, then $x^2 + y^2 \equiv 0 \bmod 3$ implies $(x/y)^2 \equiv -1 \bmod 3$, contradicting the fact that -1 is not a square modulo 3.

Thus $x = 3x_1$ and $y = 3y_1$ for natural numbers x_1, y_1. Now $3z^2 = x^2 + y^2 = 9x_1^2 + 9y_1^2$ implies $z^2 = 3(x_1^2 + y_1^2)$; since the right hand side is divisible by 3, so is the left hand side: $z = 3z_1$ for some integer $z_1 > 0$. But then $9z_1^2 = 3(x_1^2 + y_1^2)$, that is, $x_1^2 + y_1^2 = 3z_1^2$.

We have shown: given any solution (x, y, z) in natural numbers of the equation $x^2 + y^2 = 3z^2$, there is another solution (x_1, y_1, z_1) in natural numbers with $z_1 = z/3$. Repeating this argument gives yet another solution (x_2, y_2, z_2) in natural numbers with $z_2 = z_1/3 = z/9$. Eventually, this will produce a contradiction because natural numbers cannot decrease indefinitely. □

Proposition 1.23. *The equation $x^3 + 3y^3 + 9z^3 = 0$ does not have any solutions in positive integers.*

Proof. Assume that (x, y, z) is a solution in positive integers. Clearly x is divisible by 3, so $x = 3x_1$ for some positive integer x_1. But then $27x_1^3 + 3y^3 + 9z^3 = 0$, hence $9x_1^3 + y^3 + 3z^3 = 0$. Now $y = 3y_1$, and we find $3x_1^3 + 9y_1^3 + z^3 = 0$. Finally, $z = 3z_1$ for some positive integer z_1, and $x_1^3 + 3y_1^3 + 9z_1^3 = 0$.

Thus if (x, y, z) is a solution of the equation $x^3 + 3y^3 + 9z^3 = 0$ in positive integers, then so is $(\frac{x}{3}, \frac{y}{3}, \frac{z}{3})$. Repeating this argument we find that for every positive solution there is a smaller solution in positive integers: but this is nonsense, thus there is no solution in positive integers. □

Proposition 1.24. *The number $\sqrt{2}$ is irrational.*

Proof. Assume not. Then $\sqrt{2} = \frac{m}{n}$ for positive integers m, n, and squaring yields $2n^2 = m^2$. Thus $m = 2p$ is even, and we find $n^2 = 2p^2$. This shows that $n = 2q$ for some positive integer q, hence $2q^2 = p^2$. Thus if $\sqrt{2} = \frac{m}{n}$, then $\sqrt{2} = \frac{p}{q}$ with integers $p = \frac{m}{2}$ and $q = \frac{n}{2}$. Repeating this sufficiently often leads to a contradiction since no positive integer is divisible by 2 infinitely often. □

Trost's Determinant Trick

In Fermat's proof we will come across a second equation, namely $x^4 - 4y^4 = z^2$. The following "determinant trick" due to E. Trost, which is apparently completely unknown (I am not aware of a single book or article – except Trost's original publication – that mentions this idea), explains where this equation is coming from.

Assume that $x^4 + y^4 = z^2$ has a solution in integers, say (a, b, c). Then the quadratic equation

$$a^4t^2 - c^2t + b^4 = 0$$

in t has the solution $t = 1$. This can only happen if the discriminant of this equation is a rational square, that is, if $c^4 - 4(ab)^4 = w^2$.

Conversely, if (a, b, c) is an integral solution of $x^4 - 4y^4 = w^2$, then the discriminant of the quadratic $a^4t^2 - c^2t - 4b^4 = 0$ must be a square, and we find $c^4 + (2ab)^4 = w^2$, i.e., Fermat's equation for the exponent 4.

This shows that the two equations $x^4 + y^4 = z^2$ and $x^4 - 4y^4 = z^2$ are connected in some mysterious way. Actually, both are elliptic curves, and the maps from one to the other provided by Trost's trick are called 2-isogenies.[2]

The only problem with Trost's trick is that the solutions on $x^4 - 4y^4 = z^2$ coming from $x^4 + y^4 = z^2$ are *larger* than the original one; for getting the descent argument to work we have to produce *smaller* solutions.

Fermat's Proof

Fermat used descent to give a proof of

Theorem 1.25. *The Fermat equation* (1.7) *for the exponent* 4 *does not have any integral solution with $XYZ \neq 0$.*

[2]For explanations of what this means let me once more refer to the seminar on elliptic curves.

Proof. Assume that $X, Y, Z \in \mathbb{N} \setminus \{0\}$ satisfy (1.7); we may (and will) assume that these integers are pairwise coprime (otherwise we can cancel common divisors). Now we vaguely follow our solution of the Pythagorean equation: Z must be odd (if Z were even, then X and Y would have to be odd, and we get a contradiction as in the proof of Theorem 1.20).

Thus we may assume that X is odd and Y is even. Since (X^2, Y^2, Z) is a Pythagorean Triple, there exist integers m, n such that $X^2 = m^2 - n^2$, $Y^2 = 2mn$ and $Z = m^2 + n^2$. Clearly $\gcd(m, n)$ divides both X and Y, hence m and n are coprime; moreover, since X is odd, we have $1 \equiv X^2 = m^2 - n^2 \bmod 4$, which implies that m is odd and $n = 2k$ is even. Thus $(Y/2)^2 = mk$ with m and k coprime, hence $m = a^2$ and $k = b^2$, giving $X^2 = a^4 - 4b^4$.

Now we repeat this stunt: from $X^2 + 4b^4 = a^4$ we see that $(X, 2b^2, a^2)$ is a Pythagorean triple; thus $X = m_1^2 - n_1^2$, $2b^2 = 2m_1n_1$ and $a^2 = m_1^2 + n_1^2$, where m_1 and n_1 are (necessarily coprime) positive integers. From $m_1n_1 = b^2$ we deduce that $m_1 = r^2$ and $n_1 = s^2$, hence $a^2 = r^4 + s^4$, and we have found a new solution (a, r, s) to our equation $Z^2 = X^4 + Y^4$.

Since $Z = m^2 + n^2 = a^4 + 4b^4$, we find that $0 < a < Z$; this means that for every solution (X, Y, Z) in natural numbers there exists another solution with a smaller Z. This is impossible, so there can't be a nontrivial solution to the Fermat equation in the first place. □

Let us finally give a nontrivial application of Trost's determinant trick by solving the diophantine equation $x^4 - 2y^2 = z^4$. If it has a solution (a, b, c), then the discriminant of $a^4t^2 - 2b^2t - c^4$ must be a square, and we get $4b^4 + 4a^4c^4 = w^2$. Dividing through by 4 we find $a^4c^4 + b^4 = u^2$. The only rational solutions of this equation are $ac = 0$ or $b = 0$; this means that the only possible rational solutions of $x^4 - 2y^2 = z^2$ satisfy $xyz = 0$, and we have shown:

Theorem 1.26. *Every rational solution of $x^4 - 2y^2 = z^4$ satisfies $y = 0$.*

In particular, the only solutions of the equation $x^4 - 2y^2 = 1$ in integers are $(\pm 1, 0)$.

Exercises

1.1 Show that in $\mathbb{Z}$, we have $0 \mid a$ if and only if $a = 0$.

1.2 Give an explicit example of non-unique factorization in $M = \{1, 5, 9, 13, \ldots\}$.

1.3 Why does Zermelo's proof of unique factorization not work in

$$M = \{1, 5, 9, 13, \ldots\}?$$

1.4 Show that the monoid $M = \{1, 3, 5, 7, \ldots\}$ of odd natural numbers has unique factorization. (Hint: deduce it from unique factorization in $\mathbb{N}$.)

1.5 Show that the monoid $M = \{2^a 6^b \mid a, b \in \mathbb{N}_0\} = \{1, 2, 4, 6, 8, 12, 16, \ldots\}$ has unique factorization.

1.6 Let M be a monoid with $M \subseteq N$. Show that if $p \in M$ is prime in $\mathbb{N}$, then it is also prime in M. Show that the converse is not true.

1.7 Prove or disprove: if $n \mid ab$ and $n \nmid a$, then $n \mid b$.

1.8 Prove that $2 \mid n(n+1)$ for all $n \in \mathbb{N}$
a) using induction
b) directly.

1.9 Prove that $3 \mid n(n^2 - 1)$ for all $n \in \mathbb{N}$. Generalizations?

1.10 Prove that $8 \mid (n^2 - 1)$ for all odd $n \in \mathbb{N}$.

1.11 Prove or disprove: if $n \mid ab$ and $n \nmid a$, then $n \mid b$.

1.12 Show that there are arbitrary long sequences of composite numbers (Hint: observe that $2 \cdot 3 + 2$ and $2 \cdot 3 + 3$ can be seen to be composite without performing any division; generalize!)

1.13 Show that divisibility defines a *partial order* on $\mathbb{Z}$ by writing $a \leq b$ if $b \mid a$.

1.14 Show that, for integers $a, b, c, d, m \in \mathbb{Z}$ with $m > 0$, we have

- $a \equiv b \bmod m \Longrightarrow a \equiv b \bmod n$ for every $n \mid m$;
- $a \equiv b \bmod m$ and $c \equiv d \bmod m \Longrightarrow a+c \equiv b+d \bmod m$ and $ac \equiv bd \bmod m$;
- $a \equiv b \bmod m \Longrightarrow ac \equiv bc \bmod m$ for any $c \in \mathbb{Z}$.

1.15 Show that there are infinitely many primes of the form $3n - 1$.

1.16 Try to extend the above proof to the case of primes of the form $3n + 1$ (and $5n - 1$). What goes wrong?

1.17 Show that primes $p = c^2 + 2d^2$ satisfy $p = 2$ or $p \equiv 1, 3 \bmod 8$.

1.18 Show that primes $p = c^2 - 2d^2$ satisfy $p = 2$ or $p \equiv 1, 7 \bmod 8$.

1.19 Show that primes $p = c^2 + 3d^2$ satisfy $p = 3$ or $p \equiv 1 \bmod 3$.

1.20 Compute $d = \gcd(77, 105)$ and write d as a $\mathbb{Z}$-linear combination of 77 and 105.

1.21 Check the addition and multiplication table for the ring $\mathbb{Z}/3\mathbb{Z}$:

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

1.22 Compute addition and multiplication tables for the rings $\mathbb{Z}/5\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$.

1.23 Compute the multiplicative inverse of $[17]$ in $\mathbb{Z}/101\mathbb{Z}$.

1.24 Compute $\gcd(2^m - 1, 2^n - 1)$ for small values of $m, n \geq 1$ until you discover a general formula for d.

1.25 Let $U_1 = U_2 = 1$, and $U_{n+1} = U_n + U_{n-1}$ denote the Fibonacci numbers. Find a formula for $\gcd(U_m, U_n)$.

1.26 Show that the Fermat numbers $F_n = 2^{2^n} + 1$ are pairwise coprime.

1.27 Show that there are infinitely many primes of the form $p = 4n + 3$.

1.28 Show that there are infinitely many primes of the form $p = 3n + 2$.

1.29 Compute $\gcd(x^2 + 2x + 2, x^2 - x - 2)$ over $\mathbb{Z}/m\mathbb{Z}$ for $m = 2, 3, 5$ and 7, and find its Bezout representation.

1.30 Let $a, b \in \mathbb{N}$ be coprime, and let $r \in \mathbb{N}$ be a divisor of ab. Put $u = \gcd(a, r)$ and $v = \gcd(b, r)$, and show that $r = uv$.

1.31 Assume that $M_p = 2^p - 1$ is a prime. List the complete set of (positive) divisors of $N_p = 2^{p-1}M_p$, and compute their sum. Conclude that if M_p is prime, then N_p is a perfect number (a number n is called perfect if the sum of its (positive) divisors equals $2n$).

Euler later proved that every even perfect number has the form $2^{p-1}M_p$ for some Mersenne prime M_p. It is conjectured (but not known) that odd perfect numbers do not exist.

1.32 Compute the last two digits of 27^{19}.

1.33 For primes $p \in \{3, 5, 7, 11, 13\}$, compute $A \equiv (\frac{p-1}{2})! \bmod p$. Can you find a pattern? If not, compute $B \equiv A^2 \bmod p$. Formulate a conjecture.

1.34 Check which of the primes $p \in \{3, 5, 7, 11, 13\}$ can be written as $p = a^2 + b^2$ with integers $a, b \in \mathbb{N}$ (e.g. $5 = 1^2 + 2^2$). Formulate a conjecture.

1.35 For some small primes $p = 4n + 1$, compute the smallest residue (in absolute value) of $a \bmod p$, where $a = \binom{2n}{n}$. (Example: for $p = 5$, we have $n = 1$ and $\binom{2}{1} = 2 \equiv 2 \bmod 5$.) Compare with the results from the preceding Exercise. Formulate a conjecture and test it for a few more primes.

1.36 a) Given a 5-liter jar and a 3-liter jar and an unlimited supply of water, how do you measure out 4 liters exactly?
b) Can you also measure out 1, 2 and 3 liters?
c) Which quantities can you measure out if you are given a 6-liter and a 9-liter jar?
d) Formulate a general conjecture. Can you prove it (at least partially)?

1.37 Show that a number $n = d_n \dots d_1 d_0 = d_n 10^n + \dots + d_1 \cdot 10 + d_0$ satisfies the congruence $n \equiv d_n + \dots + d_1 + d_0 \bmod 9$: the residue class modulo 9 of any integer is congruent to the sum of the digits of n.

1.38 Show that a number $d_n \dots d_1 d_0 = d_n 10^n + \dots + d_1 \cdot 10 + d_0$ satisfies the congruence $n \equiv (-1)^n d_n + \dots + d_2 - d_1 + d_0 \bmod 11$.

1.39 Invent a simple method to compute the residue class of $n = d_n \dots d_1 d_0 = d_n 10^n + \dots + d_1 \cdot 10 + d_0$ modulo 7.

1.40 Compute the last digit of 7^{100}. Compute the last two digits of 3^{65}.

1.41 Prove, using the Euclidean algorithm, that $\gcd(am, an) = a \gcd(m, n)$. Hint: apply the Euclidean algorithm to the pair (m, n). What can you say about the remainders when you apply the algorithm to (am, an) instead?

1.42 Using the last exercise, prove that if $a \mid mn$ and $\gcd(a, m) = 1$, then $a \mid n$. Hint: we have $\gcd(an, mn) = n$; now observe that a divides both an and mn.

1.43 Solve the diophantine equation $x^2 + 2y^2 = z^2$.

1.44 Solve the diophantine equation $x^2 - 2y^2 = z^2$.

1.45 Solve the diophantine equation $x^2 + y^2 = 2z^2$.

1.46 Solve the diophantine equation $x^2 - y^2 = 3$.

1.47 Prove that each odd prime p can be written as a difference of squares of natural numbers ($p = y^2 - x^2$ for $x, y \in \mathbb{N}$) in a unique way.

1.48 Fermat repeatedly challenged English mathematicians by sending them problems he claimed to have solved and asking for proofs. Two of them were the following that he sent to Wallis:

- Prove that the only solution of $x^2 + 2 = y^3$ in positive integers is given by $x = 5$ and $y = 3$;
- Prove that the only solution of $x^2 + 4 = y^3$ in positive integers is given by $x = 11$ and $y = 5$.

In a letter to his English colleague Digby, Wallis called these problems trivial and useless, and mentioned a couple of problems that he claimed were of a similar nature:

- $x^2 + 12 = y^4$ has unique solution $x = 2$, $y = 2$ in integers;
- $x^4 + 9 = y^2$ has unique solution $x = 2$, $y = 5$ in integers;
- $x^3 - y^3 = 20$ has no solution in integers;
- $x^3 - y^3 = 19$ has unique solution $x = 3$, $y = 2$ in integers.

When Fermat learned about Wallis's comments, he called Wallis's problems mentioned above "amusements for a three-day arithmetician" in a letter to Digby. In fact, while Fermat's problems were hard (and maybe not even solvable using the mathematics known in his times), Wallis's claims are easy to prove. Do this.

1.49 Assume that $ab = rx^n$ for $a, b, r, x \in \mathbb{N}$ and $\gcd(a, b) = 1$. Show that there exist $u, v, y, z \in \mathbb{N}$ such that $a = uy^n$, $b = vz^n$, and $uv = r$.

1.50 Use infinite descent to prove that $\sqrt{3}$ is irrational.

1.51 Let p be a prime; show that $x^3 + py^3 + p^2z^3 = 0$ does not have a solution.

1.52 The employees of a big company are represented in a computer by 5-digit numbers. A check digit c is introduced to minimize errors. The company uses the formula $c \equiv 1 \cdot d_1 + 2 \cdot d_2 + 3 \cdot d_3 + 4 \cdot d_4 + 5 \cdot d_5 \bmod 10$ to compute the check digit of $d_1d_2d_3d_4d_5$.

1. Compute the check digits of 01716 and 01718. What do you observe? Is the formula for the check digit a good choice? Why (not)?
2. Would the formula $c \equiv d_1 + 3 \cdot d_2 + d_3 + 3 \cdot d_4 + d_5 \bmod 10$ be a better choice? What about $c \equiv d_1 + 2 \cdot d_2 + d_3 + 2 \cdot d_4 + d_5 \bmod 10$?

Chapter 2

Residue Class Rings

We have already seen that the unit group of $\mathbb{Z}$ is simply $\mathbb{Z}^\times = \{-1, +1\}$, a group of order 2. Let us now determine the unit groups of the rings of residue classes $\mathbb{Z}/m\mathbb{Z}$. Observe that a residue class u modulo m is a unit if there exists an integer v such that $[uv]_m = [1]_m$, in other words: if $uv \equiv 1 \bmod m$ for some $v \in \mathbb{Z}$.

Now we claim

Theorem 2.1. *We have* $(\mathbb{Z}/m\mathbb{Z})^\times = \{a \bmod m : \gcd(a, m) = 1\}$.

Proof. It is now that the Bezout representation begins to show its full power. If $\gcd(a, m) = 1$, then there exist integers $x, y \in \mathbb{Z}$ such that $ax + my = 1$. Reducing this equation modulo m gives $ax \equiv 1 \bmod m$, in other words: the residue class $a \bmod m$ is a unit! Not only that: the extended Euclidean algorithm gives us a method to compute the inverse elements.

To prove the converse, assume that $a \bmod m$ is a unit. Then $ac \equiv 1 \bmod m$ for some $c \in \mathbb{Z}$, so $ac = km + 1$ for some $k \in \mathbb{Z}$. But then $ac - km = 1$ shows that $\gcd(a, m) = 1$. □

If $m = p$ is a prime, the unit groups are particularly simple: we have $\gcd(a, p) = 1$ if and only if $p \nmid a$, hence $(\mathbb{Z}/p\mathbb{Z})^\times = \{1, 2, \ldots, p-1\} = \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$. But if every element $\neq 0$ of a ring has an inverse, then that ring is a field, and we have given a second proof of the following

Corollary 2.2. *If p is a prime, then the residue class ring $\mathbb{Z}/p\mathbb{Z}$ is a (finite) field.*

The field $\mathbb{Z}/p\mathbb{Z}$ is called a finite field because it has finitely many elements. As we have seen, there are finite fields with p elements for every prime p. Later we will see that there exist finite fields with $m > 1$ elements if and only if m is a prime power.

The fact that $\mathbb{Z}/p\mathbb{Z}$ is a field means that expressions like $\frac{1}{7} \bmod 11$ make sense. To compute such 'fractions', you can choose one of the following two methods:

1. Change the numerator mod 11 until the division becomes possible:

$$\frac{1}{7} \equiv \frac{12}{7} \equiv \frac{23}{7} \equiv \frac{34}{7} \equiv \frac{45}{7} \equiv \frac{56}{7} = 8 \bmod 11,$$

and in fact $7 \cdot 8 = 56 \equiv 1 \bmod 11$. This method only works well if p is small.

2. Apply the Euclidean algorithm to the pair $(7, 11)$, and compute a Bezout representation; you will find that $1 = 2 \cdot 11 - 3 \cdot 7$, and reducing mod 11 gives $1 \equiv (-3) \cdot 7 \bmod 11$, hence the multiplicative inverse of 7 mod 11 is $-3 \equiv 8 \bmod 11$.

2.1 Fermat's Little Theorem

Theorem 2.3 (Fermat's Little Theorem)**.** *If p is a prime and a an integer not divisible by p, then $a^{p-1} \equiv 1 \bmod p$.*

The following proof is due to Leibniz[1] and probably the oldest proof known for Fermat's Little Theorem. It uses binomial coefficients: these are the entries in Pascal's triangle, and they occur in the binomial theorem

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \ldots + \binom{n}{n-1} ab^{n-1} + b^n.$$

We will need two properties of $\binom{n}{k}$: first we use the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, and then we claim

Lemma 2.4. *If p is a prime, then the numbers $\binom{p}{k}$, $k = 1, 2, \ldots, p-1$, are all divisible by p.*

For example, the fifth row of Pascal's triangle is 1 5 10 10 5 1. The claim is not true if p is not a prime: the sixth row is 1 6 15 20 15 6 1, and the numbers 15 and 20 are not divisible by 6.

Proof. From $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ we see that the numerator is divisible by p while the denominator is not divisible by p unless $k = 0$ or $k = p$. Thus we conclude that $p \mid \binom{p}{k}$ for $0 < k < p$. □

Now we can give an induction proof of Fermat's Little Theorem:

Proof. We prove the equivalent (!) statement $a^p \equiv a \bmod p$ for all $a \in \mathbb{Z}$ via induction on a. The claim is clearly trivial for $a = 1$; assume it has been proved for some a; then

$$(a+1)^p = a^p + \binom{p}{1} a^{p-1} + \ldots + \binom{p}{p-1} a + 1.$$

[1] Gottfried Wilhelm von Leibniz, 1646 (Leipzig) – 1716 (Hannover).

Since the binomial coefficients are all $\equiv 0 \bmod p$ by the lemma, we find

$$(a+1)^p \equiv a^p + 1 \bmod p,$$

and by the induction assumption, $a^p \equiv a \bmod p$, so we get $(a+1)^p \equiv a+1 \bmod p$, and the induction step is established. □

There is another proof of Fermat's little theorem due to Dirichlet that works for any finite group. To see what's going on, consider $(\mathbb{Z}/5\mathbb{Z})^\times = \{[1],[2],[3],[4]\}$, where $[r]$ denotes the residue class $r \bmod 5$. If we multiply each of these classes by 3, we get

$$\begin{aligned} [1]\cdot[3] &= [3], \\ [2]\cdot[3] &= [1], \\ [3]\cdot[3] &= [4], \\ [4]\cdot[3] &= [2]; \end{aligned}$$

thus multiplying all prime residue classes mod 5 by 3 yields the same classes again, though in a different order. If we multiply these four equations together, we get $[1][2][3][4]\cdot[3]^4 = [3][1][4][2] = [1][2][3][4]$, hence $[3]^4 = [1]$, or, in other words, $3^4 \equiv 1 \bmod 5$. This can be done in general:

Second Proof of Thm. 2.3. Write $(\mathbb{Z}/p\mathbb{Z})^\times = \{[1],[2],\dots,[p-1]\}$; let a be an integer not divisible by p. If we multiply each residue class with $[a]$, we get the $p-1$ classes $[a]$, $[2a]$, …, $[(p-1)a]$:

$$\begin{aligned} [1]\cdot[a] &= [a] \\ [2]\cdot[a] &= [2a] \\ &\vdots \\ [p-1]\cdot[a] &= [(p-1)a] \end{aligned}$$

If we can show that the classes on the right hand side are all different, then they must be a permutation of the classes $[1], \dots, [p-1]$ that we started with. Taking this for granted, the products $[a]\cdot[2a]\cdots[(p-1)a] = [(p-1)!][a^{p-1}]$ and $[1]\cdot[2]\cdots[p-1] = [(p-1)!]$ must be equal (after all, the factors are just rearranged). But $(p-1)!$ is coprime to p, so we may cancel this factor, and get $[a^{p-1}] = [1]$, i.e., $a^{p-1} \equiv 1 \bmod p$.

It remains to show that the classes $[a]$, $[2a]$, …, $[(p-1)a]$ are all different. Assume therefore that $[ra] = [sa]$ for integers $1 \le r, s \le p-1$; we have to show that $r = s$. But $[ra] = [sa]$ means that $[(r-s)a] = [0]$, i.e. that $p \mid (r-s)a$. Since $p \nmid a$ by assumption, the fact that p is prime implies $p \mid (r-s)$. But $r-s$ is an integer strictly between $-p$ and p, and the only such integer is 0: thus $r = s$ as claimed. □

Here is another way to prove Fermat's Little Theorem. Consider the set of all triples xyz with $x, y, z \in \{0,1\}$. There are clearly 2^3 of them. Now we count

them differently:

$$\begin{aligned}&000\\&100 + \text{cyclic shifts}\\&110 + \text{cyclic shifts}\\&111\end{aligned}$$

Here the cyclic shifts of 100 are 010 and 001; since there are three cyclic shifts of 100 and three shifts of 110, we must have $2^6 = 2 + 3 \cdot 2$, in other words: $2^3 \equiv 2 \bmod 3$.

Now work out this proof for vectors of length 5 and show that $2^5 \equiv 2 \bmod 5$. Think about what goes wrong for exponent 6, and generalize the proof above to $2^p \equiv 2 \bmod p$. Finally, by letting the entries be from $\mathbb{Z}/p\mathbb{Z}$ instead of $\{0, 1\}$, Fermat's Little Theorem follows.

2.2 RSA

Cryptography deals with methods that allow us to transmit information safely, that is, in such a way that eavesdroppers have no chance of reading it. Simple methods for encrypting messages were known and widely used in military circles for several millenia; basically all of these codes are easy to break with computers.

An example of such a classical code is Caesar's cipher: permute the letters of the alphabet by sending X $\longmapsto$ A, Y $\longmapsto$ B, Z $\longmapsto$ C, A $\longmapsto$ D etc; the text "ET TU, BRUTE" would be encrypted as "BQ QR, YORQB". For longer texts, analyzing the frequency of letters (for given languages) makes breaking this and similar codes a breeze, in particular if you are equipped with a computer.

Another common feature of these ancient methods of encrypting messages is the following: anyone who knows the key, that is, the method with which messages are encrypted, can easily break the code by inverting the encryption. In 1976, Diffie and Hellman suggested the existence of public key cryptography: these are methods for encrypting messages that do not allow you to read encrypted messages even if you know the key. The most famous of all public key cryptosystems is called RSA after its discoverers Ramir, Shamir and Adleman (1978).

Before we explain how RSA works, we will have to generalize Fermat's Little Theorem somewhat. Let p and q be distinct odd primes. Then for any integer a coprime to pq, we have $a^{(p-1)(q-1)} \equiv 1 \bmod pq$. In fact, applying Fermat's Little Theorem twice we find

$$\begin{aligned}a^{(p-1)(q-1)} &= \left(a^{q-1}\right)^{p-1} \equiv 1 \bmod p,\\ a^{(p-1)(q-1)} &= \left(a^{p-1}\right)^{q-1} \equiv 1 \bmod q.\end{aligned}$$

Thus $p \mid (a^{(p-1)(q-1)} - 1)$ and $q \mid (a^{(p-1)(q-1)} - 1)$; since p and q are coprime, it follows from unique factorization that $pq \mid (a^{(p-1)(q-1)} - 1)$, and this implies the claim.

Here comes the simple idea behind RSA:

Now assume that Alice wants others to send her messages that cannot be read by anyone who's listening in. She picks two large prime numbers p and q (in practice, these should have about 150 digits each); she computes the product N and chooses an integer $1 < e < (p-1)(q-1)$ coprime to $(p-1)(q-1)$. Then she publishes the pair (N, e), her "public key" (for example by listing it on her homepage).

How does the encryption work? It is a simple matter to transform any text into a sequence of numbers, for example by using $a \longmapsto 01$, $b \to 02$, ..., with a couple of extra numbers for blanks, commas, etc. We may therefore assume that our message is a sequence of integers $m < n$ (if the text is longer, break it up into smaller pieces). Bob encrypts each integer m as $c \equiv m^e \bmod N$ and sends the sequence of c's to Alica (by email, say). Now Alice can decrypt the message as follows: since she knows p and q, she run the Euclidean algorithm on the pair $(e, (p-1)(q-1))$ to find integers $d, x > 0$ such that $de - x(p-1)(q-1) = 1$ (Bezout at work again). Now she takes the message c and computes $c^d \bmod N$. The result is $c^d \equiv (m^e)^d = m^{de} = m^{1+x(p-1)(q-1)} = m \cdot (m^{(p-1)(q-1)})^x \equiv m \bmod N$: thus Alice can compute the original text that Bob sent her.

Now assume that Eve is eavesdropping. Of course she knows the pair (N, e) (which is public anyway), and she also knows the message c that Bob sent to Alice. This does not suffice for decrypting the message, however, since one seems to need an inverse d of $e \bmod (p-1)(q-1)$ to do that; it is likely that one needs to know the factors of N in order to compute d.

Alice	Eve	Bob
picks two large primes p, q		
computes $N = pq$		
picks random $e \in (\mathbb{Z}/\phi(N)\mathbb{Z})^\times$		
publishes public key (N, e)	$(N, e) \longrightarrow$	
		computes $c \equiv m^e \bmod N$
	$\longleftarrow c$	sends c to Alice
solves $de \equiv 1 \bmod (p-1)(q-1)$		
computes $m \equiv c^d \bmod N$		

Figure 2.1: The RSA protocol

If Bob and Alice want to exchange messages, each of them has to pick their own key (so Alice picks primes p_A, q_A, and Bob p_B and q_B etc.)

Baby Example. The following choice of $N = 1073$ with $p = 29$ and $q = 37$ is not realistic because this number can be factored easily; its only purpose is to illustrate the method.

So assume that Alice picks the key $(N, e) = (1073, 25)$. Bob wants to send the message "miss piggy" to Alice. He starts by transforming the message into a string of integers as follows:

	m	i	s	s		p	i	g	g	y
m	13	9	19	19	27	16	9	7	7	25

Next he encrypts this sequence by computing $c \equiv m^{25} \bmod N$ for each of these m: starting with $13^{25} \equiv 671 \bmod 1073$, he finds

m	13	9	19	19	27	16	9	7	7	25
c	671	312	901	901	656	1011	312	922	922	546

Bob sends this string of c's to Alice. Knowing the prime factorization of N, Alice is able to compute the inverse of 25 mod $(p-1)(q-1)$ as follows: she multiplies $p-1=28$ and $q-1=36$ to get $(p-1)(q-1)=28\cdot 36=1008$. Then she applies the Euclidean algorithm to $(25, 1008)$ and finds $1=25\cdot 121-1008\cdot 3$, and this shows that $d=121$.

Now Alice takes the string of c's she got from Bob and decrypts them: starting with $671^{121} \equiv 13 \bmod n$ she can get back the string of m's, and hence the original message.

Remark. There is a big problem with this baby example: if we encrypt the message letter for letter, then equal letters will have equal code, and the cryptosystem can be broken (if the message is long enough) by analyzing the frequency with which each letter occurs (say in English). This problem vanishes into thin air when we use (realistic) key sizes of about 200 digits: there we encrypt the message in blocks of about 100 letters, and since the chance that any two blocks of 100 letters inside a message coincide is practically 0, an attack based on the frequency of letters will not be successful for keys of this size.

Here's a somewhat more realistic example (you will need pari to check the calculations). Pick $p = 420967$ and $q = 723451$, and put $N = pq = 304548997117$. Pick a random (odd) e mod N such as $e = 304548997117$ and check that $\gcd(e, p-1) = \gcd(e, q-1) = 1$.

Now the public key is the pair (N, e). To encrypt the message "books", Bob replaces every letter by its number in the alphabet (for longer messages, you can use 27 for a blank space etc.) and finds $m = 0215151119$. He encrypts this by computing $c \equiv m^e \equiv 95972141147 \bmod N$. Bob sends this number to Alice, who computes the Bezout representation of $1 = \gcd(e, (p-1)(q-1))$ and finds

$$(-118628497213e + 115672710009(p-1)(q-1) = 1.$$

Thus $d \equiv -118628497213 \equiv 185919355487 \bmod (p-1)(q-1)$ (we need a positive value of d). Then she computes $c^d \equiv 215151119 \bmod N$, which is, of course, the original message; translating the numbers 2, 15, ... back into letters she finds the word "books".

2.3 Pollard's $p-1$-Factorization Method

Fast Exponentiating

For applications of number theory to cryptography or factoring large integers we need an algorithm that computes $a^m \bmod n$ a lot faster than the naive al-

gorithm, which proceeds my multiplying a repeatedly to itself. Here's how it works: let $\mathrm{pow}\,(a, m, n)$ denote the residue class $[a^m]_n$. Then the following recursive procedure computes $\mathrm{pow}\,(a, m, n)$ with at most $2\log_2 m$ multiplications as opposed to $m-1$ multiplications using the naive method:

$$\mathrm{pow}\,(a, m, n) = \begin{cases} [a]_n & \text{if } m = 1, \\ \mathrm{pow}\,(a^2, \frac{m}{2}, n) & \text{if } m \text{ is even}, \\ [a]_n \cdot \mathrm{pow}\,(a^2, \frac{m-1}{2}, n) & \text{if } m > 1 \text{ is odd} \end{cases}$$

Pollard is definitely the world champion in inventing new methods for factoring integers. One of his earliest contributions were the $p-1$-method (ca. 1974), his ρ-method followed shortly after, and his latest invention is the number field sieve (which is based on ideas from algebraic number theory).

The idea behind Pollard's $p-1$-method is incredibly simple. Assume that we are given an integer N that we want to factor. Fix an integer $a > 1$ and check that $\gcd(a, N) = 1$ (should $d = \gcd(a, N)$ be not trivial, then we have already found a factor d and continue with N replaced by N/d).

Let p be a factor of N; by Fermat's Little Theorem we know that $a^{p-1} \equiv 1 \bmod p$, hence $D := \gcd(a^{p-1} - 1, N)$ has the properties $p \mid D$ and $D \mid N$. Thus D is a nontrivial factor of N unless $D = N$ (which should not happen too often).

The procedure above is not much of a factorization algorithm as long as we have to know the prime factor p beforehand. The prime p occurs at two places in the method above: first, as the modulus when computing $a^{p-1} \bmod p$. But this problem is easily taken care of because we may simply compute $a^{p-1} \bmod N$. It is more difficult to get rid of the p in the exponent: the fundamental observation is that we can replace the exponent $p-1$ above by any multiple, and D still will be divisible by p (note though that the chance that $D = N$ has become slightly larger). Does this help us? Not always; assume, however, that $p-1$ is the product of *small* primes (say of primes below a bound B that in practice can be taken to be $B = 10^5$ or $B = 10^6$, depending on the computing power of your hardware). Then it is not too hard to come up with good candidates for multiples of $p-1$: we might simply pick $k = B!$, or, in a similar vein,

$$k = \prod_i p_i^{a_i}, \quad \text{where } p_i^{a_i} \le B < p_i^{a_i+1}. \tag{2.1}$$

If we $(p-1) \mid k$, then $a^k \equiv 1 \bmod p$, hence $p \mid D = \gcd(a^k - 1, N)$.

Thus the following algorithm has a good chance of finding those factors p of N for which $p-1$ has only small prime factors:

1. `Pick` $a > 1$ `and check that` $\gcd(a, N) = 1$
2. `Choose a bound` B`, say` $B = 10^4$`,` 10^5`,` 10^6`, ...`
3. `Pick` k `as in (2.1) and compute` $D = \gcd(a^k - 1, N)$.

Note that the computation of a^k can be done modulo N; if $p \mid N$ and $(p-1) \mid k$, then $a^k \equiv 1 \bmod p$, hence $p \mid D$.

If $D = 1$, we may increase k; if $D = N$, we can reduce k and repeat the computation.

Among the record factors found by the $p-1$-method is the 37-digit factor $p = 6902861817667290192729108442204980121$ of $71^{77} - 1$ with $p - 1 = 2^3 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 401 \cdot 409 \cdot 3167 \cdot 83243 \cdot 83983 \cdot 800221 \cdot 2197387$ discovered by Dubner. A list of record factors can be found at

`http://www.users.globalnet.co.uk/~aads/Pminus1.html`

Here's a baby example: take $N = 1769$, $a = 2$ and $B = 6$. Then we compute $k = 2^2 \cdot 3 \cdot 5$ and we find $2^{60} \equiv 306 \bmod 1769$, $\gcd(305, 1769) = 61$ and $N = 29 \cdot 61$. Note that $61 - 1 = 2^2 \cdot 3 \cdot 5$, so the factor 61 was found, while $29 - 1 = 2^2 \cdot 7$ explains why 29 wasn't (although $29 < 61$).

Another large class of factorization algorithms is based on an algorithm invented by Fermat: the idea is to write an integer n as a difference of squares. If $n = x^2 - y^2$, then $n = (x-y)(x+y)$, and unless this is the trivial factorization $n = 1 \cdot n$, we have found a factor.

Another baby example: take $n = 1073$; then $\sqrt{n} = 32.756\ldots$, so we start by trying to write $n = 33^2 - y^2$. Since $33^2 - 1073 = 16$, we find $n = 33^2 - 4^2 = (33-4)(33+4) = 29 \cdot 37$. If the first attempt would have been unsuccessful, we would have tried $n = 34^2 - y^2$, etc.

In modern algorithms (continued fractions, quadratic sieve, number field sieve) the equation $N = x^2 - y^2$ is replaced by a congruence $x^2 \equiv y^2 \bmod N$: if we have such a thing, then $\gcd(x-y, N)$ has a good chance of being a nontrivial factor of N. The first algorithm above constructed such pairs (x, y) by computing the continued fraction expansion of $\sqrt{n}$ (which we have not discussed), the number field sieve produces such pairs by factoring certain elements in algebraic number fields.

2.4 The Theorem of Euler-Fermat

Consider the unit group $(\mathbb{Z}/15\mathbb{Z})^\times$ of $\mathbb{Z}/15\mathbb{Z}$. It consists of the eight residue classes [1], [2], [4], [7], [8], [11], [13], [14]. If we multiply each of these classes e.g. by [7] (or [8], [9]), then we get

$$\begin{array}{rcr@{\qquad}rcr@{\qquad}rcr}
[1]\cdot[7] &=& [7] & [1]\cdot[8] &=& [8] & [1]\cdot[9] &=& [9] \\
[2]\cdot[7] &=& [14] & [2]\cdot[8] &=& [1] & [2]\cdot[9] &=& [3] \\
[4]\cdot[7] &=& [13] & [4]\cdot[8] &=& [2] & [4]\cdot[9] &=& [6] \\
[7]\cdot[7] &=& [4] & [7]\cdot[8] &=& [11] & [7]\cdot[9] &=& [3] \\
[8]\cdot[7] &=& [11] & [8]\cdot[8] &=& [4] & [8]\cdot[9] &=& [12] \\
[11]\cdot[7] &=& [2] & [11]\cdot[8] &=& [13] & [11]\cdot[9] &=& [9] \\
[13]\cdot[7] &=& [1] & [13]\cdot[8] &=& [14] & [13]\cdot[9] &=& [12] \\
[14]\cdot[7] &=& [8] & [14]\cdot[8] &=& [7] & [14]\cdot[9] &=& [6]
\end{array}$$

As in our proof of Fermat's Little Theorem, the resulting residue classes (for multiplication by [7] and [8]) are the classes we started with in a different order.

Multiplying these equations we get

$$\prod_{(a,15)=1} [a] = \prod_{(a,15)=1} [7a] = [7]^8 \prod_{(a,15)=1} [a].$$

Since the a are coprime to 15, so is their product; thus we may cancel, and we find $[7]^8 = [1]$, or $7^8 \equiv 1 \bmod 15$. Similarly, we find $8^8 \equiv 1 \bmod 15$; for multiplication by 9, however, the classes on the right hand side differ from those on the left (they're all divisible by 3 since both 9 and 15 are), and we do *not* get $9^8 \equiv 1 \bmod 15$.

The same idea works in general. Let $m \geq 2$ be an integer, and let $\phi(m)$ denote the number of residue classes coprime to m, that is, $\varphi(m) = \#(\mathbb{Z}/m\mathbb{Z})^\times$. Then we have the following result, which is usually referred to as the Euler-Fermat Theorem: it is due to Euler, but contains Fermat's Little Theorem as a special case.

Theorem 2.5. *If a is an integer coprime to $m \geq 2$, then $a^{\varphi(m)} \equiv 1 \bmod m$.*

For $m = p$ prime, we have $\phi(p) = p - 1$, and Euler's Theorem becomes Fermat's Little Theorem.

Proof. Let $[r_i]$, $i = 1, \ldots, t = \phi(m)$, denote the residue classes in $(\mathbb{Z}/m\mathbb{Z})^\times$. Then we claim that $[ar_1]$, $\ldots$, $[ar_t]$ are pairwise distinct. In fact, assume that $[ar_i] = [ar_j]$ with $i \neq j$, that is, $ar_i \equiv ar_j \bmod m$. Since $\gcd(a, m) = 1$, we may cancel a, and get $[r_i] = [r_j]$: contradiction.

Since the classes $[ar_1]$, $\ldots$, $[ar_t]$ are all in $(\mathbb{Z}/m\mathbb{Z})^\times$ and different, and since there are only t different classes in $(\mathbb{Z}/m\mathbb{Z})^\times$, we must have $(\mathbb{Z}/m\mathbb{Z})^\times = \{[ar_1], \ldots, [ar_t]\}$. But then $\prod_{i=1}^t [r_i] = \prod_{i=1}^t [ar_i] = [a]^{\phi(m)} \prod_{i=1}^t [r_i]$. Since the $[r_i]$ are coprime to m, so is their product. Cancelling then gives $[a]^{\phi(m)} = [1]$, which proves the claim. □

2.5 Euler's Phi Function

For the application of Euler-Fermat we need a formula that allows us to compute $\phi(n)$. Let us first compute $\phi(n)$ directly for some small n. For $n = 6$, there are 6 different residue classes modulo 6; the classes [0], [2], [3] and [4] are not coprime to 6 (or, in other words, do not have a multiplicative inverse), which leaves the classes [1] and 5 as the only ones that are coprime to 6: thus $\phi(6) = 2$. The classes mod 8 coprime to 8 are [1], [3], [5], [7], hence $\phi(8) = 4$. If p is prime, then all the $p-1$ classes [1], [2], $\ldots$, $[p-1]$ are coprime to p, hence $\phi(p) = p-1$.

n	3	4	5	6	7	8	9	10	12	15
$\phi(n)$	2	2	4	2	6	4	6	4	4	8

We can easily compute $\phi(p^k)$ (Euler's phi function for prime powers): starting with all the nonzero classes [1], [2], $\ldots$, $[p^2 - 1]$ (there are $p^2 - 1$ of them) we have to eliminate those that are not coprime to p^2, that is, exactly the

multiples of p smaller than p^2: these are p, $2p$, $3p$, $\ldots$, $(p-1)p$ (note that $p \cdot p = p^2 > p^2 - 1$); since there are exactly $p-1$ of these multiples of p, there will be exactly $p^2 - 1 - (p-1) = p^2 - p = p(p-1)$ classes left: thus $\phi(p^2) = p(p-1)$.

The same method works for p^k: there are exactly $p^k - 1$ nonzero classes, namely $[1]$, $[2]$, $\ldots$, $[p^k - 1]$. The multiples of p among these classes are $[p]$, $[2p]$, $\ldots$, $p^k - p = (p^{k-1} - 1)p$, and there are exactly $p^{k-1} - 1$ of them. Thus $\phi(p^k) = p^k - 1 - (p^{k-1} - 1) = p^k - p^{k-1} = p^{k-1}(p-1)$.

We have proved

Proposition 2.6. *For primes p and integers $k \geq 1$, we have*

$$\phi(p^k) = p^{k-1}(p-1).$$

Let us now compute $\phi(pq)$ for a product of two different primes. We have $pq - 1$ nonzero residue classes $[1]$, $[2]$, $\ldots$, $[pq-1]$. The classes that have a factor in common with pq are multiples of p and multiples of q, namely $[p]$, $[2p]$, $\ldots$, $[(q-1)p$ and $[q]$, $[2q]$, $\ldots$, $[(p-1)q]$. Since there are no multiples of p that are multiples of q (like $[0]$, $[pq]$, etc) among these, there will be exactly $pq - 1 - (p-1) - (q-1) = pq - p - q + 1 = (p-1)(q-1)$ classes left after eliminating multiples of p or q. Thus $\phi(pq) = (p-1)(q-1) = \phi(p)\phi(q)$.

The general result is

Proposition 2.7. *If m and n are coprime integers, then $\phi(mn) = \phi(m)\phi(n)$.*

Before we turn to the proof, let's see how it works in a specific example like $m = 5$ and $n = 3$. What we'll do is take a residue class modulo 15 and coprime to 15, and map it to a pair of residue classes mod 3 and mod 5:

a mod 15	1	2	4	7	8	11	13	14
a mod 3	1	2	1	1	2	2	1	2
a mod 5	1	2	4	2	3	1	3	4

Thus we have the following pairs of residue classes modulo 3 and 5: $(1,1)$, $(1,2)$, $(1,3)$, $(1,4)$ and $(2,1)$, $(2,2)$, $(2,3)$, $(2,4)$. In particular, there are $\phi(5) = 4$ pairs with $a \equiv 1 \bmod 3$ and 4 pairs with $a \equiv 2 \bmod 3$.

Proof of Prop. 2.7. We have to find a map sending a residue class modulo mn to two residue classes modulo m and n. Let's try

$$\psi : (\mathbb{Z}/mn\mathbb{Z})^\times \longrightarrow (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times : [a]_{mn} \longmapsto ([a]_m, [a]_n).$$

All that's left to do is check that it works. First observe that $\gcd(ab, n) = 1$ if and only if $\gcd(a, n) = \gcd(b, n) = 1$.

Surjectivity: We have to show that, given residue classes $[r]_m$ and $[s]_n$, there exists a residue class $[a]_{mn}$ such that $[a]_m = [r]_m$ and $[a]_n = [s]_n$. At this point, Bezout comes in again: since $\gcd(m, n) = 1$, there exist $x, y \in \mathbb{Z}$ such that $1 = mx + ny$. Now put $a = ryn + sxm$: then $a = ryn + sxm \equiv ryn \equiv$

$1 \bmod m$ since $yn \equiv 1 \bmod m$ from the Bezout representation, and similarly $a = ryn + sxm \equiv sxm \equiv s \bmod n$.

Injectivity: Assume that there are residue classes $[a]_{mn}$ and $[b]_{mn}$ such that $[a]_m = [b]_m$ and $[a]_n = [b]_n$. Then $m \mid (b-a)$ and $n \mid (b-a)$, and since $\gcd(m,n) = 1$, this implies that $[a]_{mn} = [b]_{mn}$ and proves the injectivity of ϕ. □

Here is how one could come up with the application of Bezout in the above proof. Given coprime residue classes $r \bmod m$ and $s \bmod m$, we want a formula for computing an integer a such that $a \equiv r \bmod m$ and $a \equiv s \bmod n$. The first idea is to see whether a can be written as a linear combination of r and s, that is, to look for integers x, y such that $a = xr + ys$. Reduction modulo m gives

$$r \equiv a = xr + ys \bmod m. \tag{2.2}$$

The simplest way to achieve this is by taking $x = 1$ and $y = 0$. But observe that we also need

$$s \equiv a \equiv xr + ys \bmod n. \tag{2.3}$$

Thus we need more leeway. The right idea is to observe that (2.2) will be satisfied if only $x \equiv 1 \bmod m$, $y \equiv 0 \bmod m$. Similarly, (2.3) will be satisfied if $x \equiv 0 \bmod n$ and $y \equiv 1 \bmod n$.

Is it possible to satisfy these four congruences simultaneously? Let's see: $x \equiv 0 \bmod n$ and $y \equiv 0 \bmod m$ mean $x = an$ and $y = bm$ for some $a, b \in \mathbb{Z}$. The two other congruences boil down to $x = an \equiv 1 \bmod m$ and $y = bm \equiv 1 \bmod n$. But these are both solvable since $\gcd(m,n) = 1$, so n has an inverse a modulo m, and m has an inverse b modulo n. Inverses can be computed using Bezout, and collecting everything we now can see where the formulas in the above proof were coming from.

Combining the formulas for Euler's phi function for prime powers and for products of coprime integers, we now find that an integer

$$m = p_1^{a_1} \cdots p_r^{a_r}$$

has exactly

$$\begin{aligned}\phi(m) &= (p_1 - 1)p_1^{a_1 - 1} \cdots (p_r - 1)p_r^{a_r - 1} \\ &= p_1^{a_1} \cdots p_r^{a_r} \cdot \frac{p_1 - 1}{p_1} \cdots \frac{p_r - 1}{p_r} \\ &= m\Big(1 - \frac{1}{p_1}\Big) \cdots \Big(1 - \frac{1}{p_r}\Big)\end{aligned}$$

residue classes coprime to m.

Chinese Remainder Theorem

In the proof of the multiplicativity of Euler's phi function we have shown that, given a system of congruences

$$x \equiv a \bmod m$$
$$y \equiv b \bmod n$$

can always be solved if m and n are coprime. This result, or rather its generalization to system of arbitrarily many such congruences, is called the Chinese Remainder Theorem.

The Chinese Remainder Theorem is often used to speed up calculations. Consider e.g. the problem of decrypting messages in the RSA system: given some $c \mod N$ for $N = pq$, the decryption requires computing $m \equiv c^d \bmod N$, where d typically has hundreds of digits. If you want fast decryption on a system with small computing power, such as a cell phone, then it would be nice if we could use a very small decryption exponent d, such as $d = 3$. Picking $d =< 10$, say, would be quite foolish, however, since someone who knows c and N could simply try all small values of d and just check whether the decrypted message makes sense.

Now this is where the Chinese Remainder Theorem comes in. Assume that $3 \nmid (p-1)$ and $5 \nmid (q-1)$, for example, and solve the system of congruences

$$d \equiv 3 \bmod p-1, \qquad d \equiv 5 \bmod q-1.$$

Then d will be large, in general about the size of $(p-1)(q-1)$; yet decryption can be performed in a very efficient way: First, find a Bezout representation $1 = ap + bq$. Then, for each encrypted message c, compute $m_1 \equiv c^3 \bmod p$ and $m_2 \equiv c^5 \bmod q$, which is very fast since you need only $2+3 = 5$ multiplications modulo p or q. Finally, put $M \equiv bqm_1 + apm_2 \bmod N$: then

$$M \equiv bqm_1 \equiv m_1 \equiv c^3 \equiv c^d \bmod p,$$
$$M \equiv apm_2 \equiv m_2 \equiv c^5 \equiv c^d \bmod q,$$

hence $M \equiv c^d \equiv m \bmod N$.

Application to Secret Sharing

Consider the following problem: you have a group of 10 people who are responsible for taking care of a bank account. If you give each of these access to the account, one of them might transfer money to his own account. If you demand that a transfer is only possible if all 10 of these people agree, then you will get nothing done because it rarely happens that all 10 are around at the same time.

What you would like to have is a procedure that allows any three of them to access the acount. The problem is to split up the password in such a way that any three of them can reconstruct it.

I will present two solutions to this problem; the first one uses linear algebra, the second the Chinese Remainder Theorem.

Linear Algebra.

In fact, assume that your password for the day is a number N, say $N = 3141$. Now you compute ten linear equations in three unknowns that all have solutions $(x, y, z) = (N, *, *)$, where the $*$ are arbitrary numbers, and you make sure that any three of these equations are linear independent. You hand out one equation to each person. As soon as three of them are together, they can solve the linear system and find $x = N$.

Here's an example: Take $N = 3141$ as above, and compute linear equations $ax + by + cz = d$ with the property that $(x, y, z) = (3141, 2011, 473)$ is a solution

person	a	b	c	d
1	3	2	−15	6350
2	−2	4	−7	−1549
3	3	−7	2	−3708
⋮				⋮
10	6	−9	−1	274

(I picked a, b, c "at random" and computed d from the solution (x, y, z)). Now each of these equations is given to exactly one person. If 1, 3 and 10 meet, they set up the linear system

$$\left(\begin{array}{ccc|c} 3 & 2 & -15 & 6350 \\ 3 & -7 & 2 & -3708 \\ 6 & -9 & -1 & 274 \end{array}\right)$$

Performing the usual row operations we get

$$\left(\begin{array}{ccc|c} 3 & 2 & -15 & 6350 \\ 0 & -9 & 17 & -10058 \\ 0 & -13 & 29 & -12426 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 3 & 2 & -15 & 6350 \\ 0 & 1 & -17/9 & 10058/9 \\ 0 & 1 & -29/13 & 12426/13 \end{array}\right) \longrightarrow$$

$$\left(\begin{array}{ccc|c} 3 & 2 & -15 & 6350 \\ 0 & 1 & -17/9 & 10058/9 \\ 0 & 0 & -40/117 & -18920/117 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 3 & 2 & -15 & 6350 \\ 0 & 1 & -17/9 & 10058/9 \\ 0 & 0 & 1 & 473 \end{array}\right) \longrightarrow$$

$$\left(\begin{array}{ccc|c} 3 & 2 & -15 & 6350 \\ 0 & 1 & 0 & 2011 \\ 0 & 0 & 1 & 473 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3141 \\ 0 & 1 & 0 & 2011 \\ 0 & 0 & 1 & 473 \end{array}\right)$$

and now they can read off the password $N = 3141$.

What happens when only two out of the ten people meet? For example, 1 and 2 can set up the system

$$\left(\begin{array}{ccc|c} 3 & 2 & -15 & 6350 \\ -2 & 4 & -7 & -1549 \end{array}\right)$$

Adding the second line to the first and then proceeding as usual gives

$$\left(\begin{array}{ccc|c} 1 & 6 & -22 & 4801 \\ -2 & 4 & -7 & -1549 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 6 & -22 & 4801 \\ 0 & 16 & -51 & 8053 \end{array}\right)$$

Putting $x_3 = s$, the general solution is given by $x_2 = \frac{1}{16}(8053 + 51s)$ and $x_1 = 4801 + 22s - 6x_2 = \frac{1}{8}(23s + 14249)$.

If one of these two has a background in number theory, then he will see that if x_1 is an integer, s must be chosen in the form $8n + 1$. For $n = 1, 2, 3 \ldots$ we now get the following possible values of N:

n	1	2	...	59
s	9	17	...	473
N	1807	1830	...	3141

What this means is that there are way too many possibilities for the right answer (what's more, n might be negative as well); guessing will not help, in particular if the password in real-life examples is a lot bigger.

Chinese Remainder Theorem.

Instead of splitting up the information by encoding it into linear equations, one can also use the Chinese Remainder Theorem to achieve this goal. Given a password P to be distributed among k people, find k pairwise coprime integers N_i slightly bigger than $\sqrt[3]{P}$, and compute $n_i \equiv P \bmod N_i$. Then hand out the k pairs (n_i, N_i). As soon as three employees get together, they can use CRT to solve the system of congruences $P \equiv n_i \bmod N_i$ for three distinct indices i, j, k, and they will find $P \bmod N_i N_j N_k$. Since $N_i N_j N_k > P$, this congruence determines P: all you have to do is pick the smallest positive solution.

The Abstract Version

There is more to the bijection

$$\psi : (\mathbb{Z}/mn\mathbb{Z})^\times \longrightarrow (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times : [a]_{mn} \longmapsto ([a]_m, [a]_n)$$

constructed above than meets the eye: we claim that ψ induces an isomorphism $(\mathbb{Z}/mn\mathbb{Z})^\times \longrightarrow (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$.

A homomorphism between groups $(G, \circ)$ and $(H, *)$ is a map $f : G \longrightarrow H$ that respects the group laws in the sense that we have $f(g \circ g') = f(g) * f(g')$. Here are some examples:

1. the exponential function is a homomorphism $\exp : (\mathbb{R}, +) \longrightarrow (\mathbb{R}_{>0}, \cdot)$ because $\exp(a + b) = \exp(a) \exp(b)$.

2. the logarithm is a homomorphism $\log : (\mathbb{R}_{>0}, \cdot) \longrightarrow (\mathbb{R}, +)$ because $\log ab = \log a + \log b$. Note that exp and log are inverse maps of each other.

3. The set C^∞ of all infinitely often differentiable functions $(0, 1) \longrightarrow \mathbb{R}$ is an additive group, and $\frac{d}{dx} : C^\infty \longrightarrow C^\infty$ is a homomorphism because $(f + g)' = f' + g'$.

4. If $f : V \longrightarrow W$ is a linear map between K-vector spaces V and W, then f is also a homomorphism between the additive groups $(V, +)$ and $(W, +)$.

5. The map $\psi : (\mathbb{Z}/mn\mathbb{Z})^\times \longrightarrow (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$ is a homomorphism. In fact we have

$$\begin{aligned}\psi([ab]_{mn}) &= ([ab]_m, [ab]n),\\ \psi([a]_{mn}) &= ([a]_m, [a]n),\\ \psi([b]_{mn}) &= ([b]_m, [b]n),\end{aligned}$$

and by the group law in direct products we see that

$$\psi([ab]_{mn}) = \psi([a]_{mn})\psi([b]_{mn}).$$

If $(G, \circ)$ and $(H, *)$ are groups, then the cartesian product $G \times H$ can be given a group structure by defining $(g, h)(g', h') = (g \circ g', h \circ h')$. Checking the axioms is straightforward. Also, if $G \times H$ is abelian if and only if G and H are.

Observe that if $f : G \longrightarrow H$ is a homomorphism between additively written groups, then $f(0) = 0$ and $f(-g) = -f(g)$. This follows easily from the axioms.

Since we have already seen that ψ is bijective, we can conclude that it is an isomorphism. Note that for any bijective homomorphism $f : G \longrightarrow H$ there exists a homomorphism $g : H \longrightarrow G$ such that $f \circ g$ and $g \circ f$ are the identity maps on H and G, respectively.

We can play this game also with rings: a map from a ring R to some ring S is called a ring homomorphism if $f(r + r') = f(r) + f(r')$, $f(rr') = f(r)f(r')$, and $f(1) = 1$. It is then easy to show that ψ actually induces a ring isomorphism $\mathbb{Z}/mn\mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$: this is the abstract formulation of the Chinese Remainder Theorem.

2.6 The Order of Residue Classes

Assume that we are given an integer m and an integer a coprime to m. The smallest exponent $n > 0$ such that $a^n \equiv 1 \bmod m$ is called the order of a mod m; we write $n = \operatorname{ord}_m(a)$. Note that we always have $\operatorname{ord}_m(1) = 1$. Here's a table for the orders of elements in $(\mathbb{Z}/7\mathbb{Z})^\times$:

$a \bmod 7$	1	2	3	4	5	6
$\operatorname{ord}_7(a)$	1	3	6	3	6	2

If $m = p$ is prime, then Fermat's Little Theorem gives us $a^{p-1} \equiv 1 \bmod p$, i.e., the order of a mod p is at most $p-1$. In general, the order of a is not $p-1$; it is, however, always a divisor of $p - 1$ (as the table above suggested):

Proposition 2.8. *Given a prime p and an integer a coprime to p, let n denote the order of a modulo p. If m is any integer such that $a^m \equiv 1 \bmod p$, then $n \mid m$. In particular, n divides $p - 1$.*

Proof. Write $d = \gcd(n, m)$ and $d = nx + my$; then $a^d = a^{nx+my} \equiv 1 \bmod p$ since $a^n \equiv a^m \equiv 1 \bmod p$. The minimality of n implies that $n \leq d$, but then $d \mid n$ shows that we must have $d = n$, hence $n \mid m$. □

Here comes a pretty application to prime divisors of Mersenne and Fermat numbers.

Corollary 2.9. *If p is an odd prime and if $q \mid M_p$, then $q \equiv 1 \bmod 2p$.*

Proof. It suffices to prove this for prime values of q (why?). So assume that $q \mid 2^p - 1$; then $2^p \equiv 1 \bmod q$. By Proposition 2.8, the order of 2 mod p divides p, and since p is prime, we find that $p = \operatorname{ord}_p(a)$.

On the other hand, we also have $2^{q-1} \equiv 1 \bmod p$ by Fermat's little theorem, so Proposition 2.8 gives $p \mid (q-1)$, and this proves the claim because we clearly have $q \equiv 1 \bmod 2$. □

Example: $M_{11} = 2047 = 23 \cdot 89$.

Fermat numbers are integers $F_n = 2^{2^n} + 1$ (thus $F_1 = 5$, $F_2 = 17$, $F_3 = 257$, $F_4 = 65537$, ...), and Fermat conjectured (and once even seemed to claim he had a proof) that these integers are all primes. These integers became much more interesting when Gauss succeeded in proving that a regular p-gon, p an odd prime, can be constructed with ruler and compass if p is a Fermat prime. Gauss also stated that he had proved the converse, namely that if a regular p-gon can be constructed by ruler and compass, then p is a Fermat prime, but the first (almost) complete proof was given by Pièrre Wantzel.[2]

Corollary 2.10. *If q divides F_n, then $q \equiv 1 \bmod 2^{n+1}$.*

Proof. It is sufficient to prove this for prime divisors q. Assume that $q \mid F_n$; then $2^{2^n} + 1 \equiv 1 \bmod q$, hence $2^{2^n} \equiv -1 \bmod q$ and $2^{2^{n+1}} \equiv 1 \bmod q$. We claim that actually $2^{n+1} = \operatorname{ord}_q(2)$: in fact, Proposition 2.8 says that the order divides 2^{n+1}, hence is a power of 2. But 2^{n+1} is clearly the smallest power of 2 that does it.

On the other hand, $2^{q-1} \equiv 1 \bmod q$ by Fermat's Little Theorem, and Proposition 2.8 gives $2^{n+1} \mid (q-1)$, which proves the claim. □

Actually we can improve this: since $F_n \equiv 1 \bmod 8$ for $n \geq 2$ we know that $(2/F_n) = +1$, hence $2^{(q-1)/2} \equiv 1 \bmod q$, and now $2^{n+1} \mid \frac{q-1}{2}$, which shows

Corollary 2.11. *If q divides F_n, then $q \equiv 1 \bmod 2^{n+2}$.*

In particular, the possible prime divisors of $F_5 = 4294967297$ are of the form $q = 128m + 1$. After a few trial divisions one finds $F_5 = 641 \cdot 6700417$. This is how Euler disproved Fermat's conjecture. Today we know the prime factorization of F_n for all $n \leq 11$, we know that F_n is composite for $5 \leq n \leq 30$ (and several larger values up to $n = 382447$), and we don't know any factors for $n = 14, 20, 22$ and 24. See
`http://www.prothsearch.net/fermat.html`
for more.

[2] Pièrre Wantzel, 1814 (Paris) – 1848 (Paris).

2.7 Existence of Primitive Roots

Consider the multiplicative group $(\mathbb{Z}/m\mathbb{Z})^\times$ of the ring $\mathbb{Z}/m\mathbb{Z}$. Since $a^{\phi(m)} \equiv 1 \bmod m$ for all $a \in (\mathbb{Z}/m\mathbb{Z})^\times$, the order of every such a divides $\phi(m)$. It is natural to ask whether this is best possible, i.e., whether there exists an element $g \in (\mathbb{Z}/m\mathbb{Z})^\times$ with maximal possible order $\operatorname{ord}_m(g) = \phi(m)$.

As the example $m = 8$ shows, where $\phi(m) = 4$, but $\operatorname{ord}_8(a) \leq 2$ for all $a \in (\mathbb{Z}/8\mathbb{Z})^\times$ (remember that $a^2 \equiv 1 \bmod 8$ for all odd integers a), this is not true. On the other hand, $\operatorname{ord}_3(2) = 2 = \phi(3)$, $\operatorname{ord}_5(2) = 4 = \phi(5)$, and $\operatorname{ord}_7(3) = 6 = \phi(7)$ show that elements of maximal possible order exist for all small primes.

Elements $g \in (\mathbb{Z}/m\mathbb{Z})^\times$ with $\operatorname{ord}_m(g) = \phi(m)$ are called *primitive roots* modulo m. A slightly different characterization of primitive roots is the following:

Lemma 2.12. *An element $g \in (\mathbb{Z}/m\mathbb{Z})^\times$ is a primitive roots if and only if every $a \in (\mathbb{Z}/m\mathbb{Z})^\times$ can be written in the form $a \equiv g^k \bmod m$ for some integer k.*

Proof. Exercise. □

We will now prove that there always exist primitive roots modulo pimes p. In our proof, which is due to Gauss, we will need the following

Lemma 2.13. *For every $n \in \mathbb{N}$ we have $\sum_{d|n} \phi(d) = n$.*

In fact, for $n = 6$ this says $\phi(1) + \phi(2) + \phi(3) + \phi(6) = 1 + 1 + 2 + 2 = 6$.

Proof. Consider the fractions $\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$. For some $d \mid n$, how many of these fractions have denominator d when written in lowest terms?

Clearly there will be $\phi(n)$ fractions with denominator n since these are exactly the $\frac{k}{n}$ with $\gcd(k, n) = 1$.

Now assume that $n = dm$; the fraction $\frac{k}{n}$ will have denominator d if and only if $k = mt$ and $\gcd(t, d) = 1$. Clearly there are $\phi(d)$ such fractions.

Thus among the n fractions, for each $d \mid n$ there are $\phi(d)$ fractions with denominator d, hence $n = \sum_{d|n} \phi(n)$. □

Gauss's idea for proving the existence of primitive roots can be expressed as follows:

Theorem 2.14. *Let G be a finite group. Assume that, for every divisor d of $n = \#G$, the equation $x^d = 1$ has at most d solutions. Then G is cyclic.*

Note that a finite group is called cyclicif there is an element $g \in G$ such that every element of g is some power of g.

Proof. Assume that $d \mid n$, and let $\psi(d)$ denote the number of elements in G with order d (thus for $G = (\mathbb{Z}/5\mathbb{Z})^\times$, we have $\psi(1) = 1$, $\psi(2) = 1$, and $\psi(4) = 2$). If $\psi(d) \neq 0$, then there is an element $g \in G$ of order d, and then $1, g, g^2, \ldots, g^{d-1}$

are distinct solutions of the equation $x^d = 1$ in G. By assumption, there are at most that many solutions, hence these are all solutions of $x^d = 1$.

Let us now determine the order of g^k for $0 \leq k < d$. We claim that g^k has order $d/\gcd(d,k)$. In fact, $(g^k)^{d/\gcd(d,k)} = (g^{k/\gcd(d,k)})^d = 1$, so the order of g^k divides $d/\gcd(d,k)$. On the other hand, from $1 = (g^k)^m = g^{km}$ we deduce that $d \mid km$, since d is the order of g. Dividing through by $\gcd(d,k)$ gives $\frac{d}{\gcd(d,k)} \mid \frac{k}{\gcd(d,k)}m$. But since $\frac{d}{\gcd(d,k)}$ and $\frac{k}{\gcd(d,k)}$ are coprime (we have divided out the common factors), this implies that $\frac{d}{\gcd(d,k)} \mid m$, which proves our claim.

Thus if g has order d, then there are exactly $\phi(d)$ elements of order d in G, namely the g^k with $\gcd(d,k) = 1$. In other words: we have $\psi(d) = 0$ or $\psi(d) = \phi(n)$.

Now clearly every element of g has some order, and this order divides $n = \#G$, hence $n = \sum_{d|n} \psi(d)$. Next $\psi(d) \leq \phi(d)$ implies that $n = \sum_{d|n} \psi(d) \leq \sum_{d|n} \phi(d) = n$, where we have used that $\sum_{d|n} \phi(d) = n$. We now see that we must have equality in $\sum_{d|n} \psi(d) \leq \sum_{d|n} \phi(d)$. But this happens if and only if $\psi(d) = \phi(d)$ for every $d \mid n$, and in particular there exists an element of order n since $\psi(n) = \phi(n) \geq 1$. □

Note that our proof also showed:

Corollary 2.15. *There exist exactly $\phi(p-1)$ primitive roots modulo primes p.*

In particular, there are $2 = \phi(6)$ primitive roots modulo 7 (namely 3 and 5), and there are $4 = \phi(10)$ primitive roots modulo 11.

Exercises

2.1 Let $n \equiv 7 \bmod 8$ be a positive integer. Show that n cannot be written as a sum of three squares.

2.2 Let $n \equiv 4 \bmod 9$ be an integer. Show that n cannot be written as a sum of four cubes.

2.3 Compute the addition and multiplication tables for the ring $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, and compare the result to those for $\mathbb{Z}/4\mathbb{Z}$.

2.4 Do the same exercise for the rings $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$.

2.5 Find all integers with $\phi(m) = 6$.

2.6 Show that m is prime if and only if $\phi(m) = m - 1$.

2.7 Solve the system of congruences

$$x \equiv 12 \bmod 13,$$
$$x \equiv 7 \bmod 19.$$

2.8 Determine the orders of all elements in $(\mathbb{Z}/15\mathbb{Z})^\times$.

2.9 Compute $\phi(180)$.

2.10 Prove or disprove: $\phi(n^2) = n(n-1)$ for every integer $n \geq 2$.

2.11 (Form the first round of the German mathematical olympiad 2006; it is the traditionally "easy" first problem).

Find two consecutive integers with the property that the sum of their digits is each divisible by 2006.

2.12 Let $p \neq 2, 5$ be a prime; let r denote the period length of the decimal expansion of $\frac{1}{p}$.

1. compute the period length r for the primes $p = 7$, 11, 13, and 37. Observations?
2. Let $x = \frac{1}{p}$. Compare the decimal expansion of $10^r x$ with that of x and conclude that $10^r x - x$ is an integer.
3. Show that $10^r \equiv 1 \bmod p$.
4. Prove that r divides $p - 1$.

2.13 Is 2 a primitive root modulo 17?

2.14 Find all 4 primitive roots modulo 11.

2.15 Show that 4 is not a primitive root for any odd prime p.

2.16 Is there a primitive root modulo 12? If yes, give one; if no, why not?

2.17 Consider the password $P = 768462011$, and consider the moduli $n_1 = 919$, $n_2 = 929$, $n_3 = 937$, $n_4 = 941$, and $n_5 = 947$ (these are all primes $> \sqrt[3]{P}$).

1. Compute $p_i \equiv P \bmod n_i$ with $0 < p_i < n_i$ for $i = 1, \ldots, 5$.
2. Solve the system $x \equiv p_i \bmod n_i$ for $i = 1, 2, 3$ using the Chinese remainder theorem (find the Bezout presentations and follow the notes) and check that the smallest positive solution is $x = P$ (feel free to use pari – this is what it's good for).
3. Do the same for the system $x \equiv p_i \bmod n_i$ for $i = 2, 3, 5$.

2.18

1. Compute the order $\operatorname{ord}_p(10)$ of 10 mod p for the primes $p = 3$, 7, and 11, and complete the table below.

p	3	7	11	13	17	19
$\operatorname{ord}_p(10)$				6	16	18

2. For which of these primes is 2 a primitive root?

3. The following table gives the decimal expansion of $1/p$ for the primes $p \nmid 10$ up to 19 (the bar denotes periodicity):

$$\begin{aligned} 1/3 &= .\overline{3} \\ 1/7 &= .\overline{142857} \\ 1/11 &= .\overline{09} \\ 1/13 &= .\overline{076923} \\ 1/17 &= .\overline{0588235294117647} \\ 1/19 &= .\overline{052631578947368421} \end{aligned}$$

Compare the length of the period with your table above. What is the pattern?

* Prove your conjecture.

2.19 Assume that $2^{n-1} \equiv 1 \bmod n$ for some odd integer n. Put $N = 2^n - 1$ and show that $2^N \equiv 1 \bmod N$.

2.20 A composite integer n is called a 2-pseudoprime if $2^n \equiv 1 \bmod p$. Show that there are infinitely many 2-pseudoprimes (Hint: $2^{11} - 1 = 23 \cdot 89$).

2.21 Let $p \equiv 1 \bmod 3$ be a prime, and let g be a primitive root mod p. Show that, for $x = g^{(p-1)/3}$, we have $x^3 \equiv 1 \bmod p$ and $x \not\equiv 1 \bmod p$.

2.22 Let $p \equiv 1 \bmod 3$ be a prime, and let x be as in the preceding exercise. Show that $(2x+1)^2 \equiv -3 \bmod p$.

2.23 Let $p \equiv 1 \bmod 4$ be a prime, and let g be a primitive root mod p. Show that, for $x = g^{(p-1)/4}$, we have $x^2 \equiv -1 \bmod p$.

2.24 Let n be an odd integer, and let $p \equiv 1 \bmod n$ be an odd prime. Show that the congruence $x^n \equiv 1 \bmod p$ has a solution $x \not\equiv 1 \bmod p$.

Chapter 3

Quadratic Residues

3.1 Quadratic Residues

Let b be an integer; an integer a coprime to b is called a quadratic residue modulo b if $a \equiv x^2 \bmod b$ for some integer x, and a quadratic nonresidue modulo b otherwise. The quadratic residues modulo 7 are 1, $2 \equiv 3^2$ and 4, whereas 2, 5 and 6 are quadratic nonresidues modulo 7. We have already proved that -1 is a quadratic residue modulo p for primes $p \equiv 1 \bmod 4$, and a quadratic nonresidue for primes $p \equiv 3 \bmod 4$.

Proofs without Primitive Roots

Lemma 3.1. *There are exactly $\frac{p-1}{2}$ quadratic residues modulo an odd prime p, namely the squares of the integers $1, 2, \ldots, \frac{p-1}{2}$.*

Proof. Clearly the residue classes $1^2, 2^2, \ldots, k^2 \bmod p$, where $p = 2k+1$, are quadratic residues modulo p. We claim that they are pairwise distinct. In fact, assume that $i^2 \equiv j^2 \bmod p$ for $1 \leq i, j, \leq k$. Then $p \mid (i^2 - j^2) = (i-j)(i+j)$. Since $2 \leq i+j \leq p-1$ we have $p \nmid (i+j)$; since p is prime, this implies $p \mid (i-j)$. Now $-k < i - j < k$, and since the only integer in this interval that is divisible by p is 0, we conclude that $i = j$.

Actually what we have shown is that the function $f(x) = x^2$ is injective as a function of $[1,k] \longrightarrow \mathbb{Z}/p\mathbb{Z}$. In particular, there are at least k quadratic residues. Actually, there aren't any others: if $a \equiv x^2 \bmod p$, then we can reduce $x \bmod p$ in such a way that $-k \leq x \leq k$, and replacing x by $-x$ if necessary we see that $a \equiv x^2 \bmod p$ for some $x \in [1,k]$. □

Since there are $p-1$ nonzero residue classes modulo p and $\frac{p-1}{2}$ of them are squares, this implies that there exist exactly $p - 1 - \frac{p-1}{2} = \frac{p-1}{2}$ quadratic nonresidues modulo p. These can be represented as follows:

Lemma 3.2. *Let p be an odd prime and n some quadratic nonresidue. Then the $k = \frac{p-1}{2}$ quadratic nonresidues are given by $n \cdot 1^2, n \cdot 2^2, \ldots, n \cdot k^2$.*

Proof. None of these numbers is a quadratic residue: in fact, $n \cdot r^2 \equiv s^2 \bmod p$ implies that $n \equiv (sr^{-1})^2 \bmod p$ is a square, contradicting our assumption.

Moreover, these numbers are pairwise distinct: $n \cdot r^2 \equiv n \cdot s^2 \bmod p$ implies $r^2 \equiv s^2 \bmod p$, which by the proof of the preceding lemma is only possible if $r = s$.

Since there exist exactly k quadratic nonresidues, the list above must be complete. □

Let us also introduce the following notation: we write

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & \text{if } a \text{ is a quadratic residue modulo } p, \\ -1 & \text{if } a \text{ is a quadratic nonresidue modulo } p. \end{cases}$$

The symbol $(\frac{a}{p})$ is called the quadratic Legendre symbol. Here are its most basic properties:

Proposition 3.3. *Let p be an odd prime; then*

1. *$a \equiv b \bmod p$ implies $(\frac{a}{p}) = (\frac{b}{p})$;*
2. *the Legendre symbol is multiplicative: $(\frac{a}{p})(\frac{b}{p}) = (\frac{ab}{p})$ for $a, b \in \mathbb{Z}$ coprime to p.*

Proof. The first property is clear: If $a \equiv x^2 \bmod p$ and $a \equiv b \bmod p$, then $b \equiv x^2 \bmod p$ and vice versa.

As for the second claim, there are several cases.

1. $(\frac{a}{p}) = (\frac{b}{p}) = +1$: then $a \equiv r^2 \bmod p$ and $b \equiv s^2 \bmod p$, hence $ab \equiv (rs)^2 \bmod p$ and therefore $(\frac{ab}{p}) = 1$.
2. $(\frac{a}{p}) = +1$, $(\frac{b}{p}) = -1$: then $a \equiv r^2 \bmod p$ and $b \equiv n \cdot s^2 \bmod p$, hence $ab \equiv n \cdot (rs)^2 \bmod p$ and therefore $(\frac{ab}{p}) = -1$.
3. $(\frac{a}{p}) = -1$, $(\frac{b}{p}) = +1$: same as above.
4. $(\frac{a}{p}) = (\frac{b}{p}) = -1$: then $a \equiv nr^2 \bmod p$ and $b \equiv ns^2 \bmod p$, hence $ab \equiv (nrs)^2 \bmod p$ and therefore $(\frac{ab}{p}) = 1$.

□

How can we tell whether a given integer is a quadratic residue or not? The following result does not seem to be very useful at first:

Proposition 3.4 (Euler's Criterion)**.** *We have*

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

Proof. Euler's proof proceeds as follows: let $p = 2m + 1$ be prime. If a is a square, then $a^m \equiv 1 \bmod p$ by Fermat's little theorem. Conversely, assume that $a^m \equiv 1 \bmod p$, and let $a_1, \ldots, a_m$ denote the quadratic residues modulo p. If a was a quadratic nonresidue modulo p, then $a_1 a, \ldots, a_m a$ would be the quadratic nonresidues. But then $x^m \equiv 1 \bmod p$ for all the quadratic residues $x = a_i$ and all the nonresidues $x = a_i a$ as well since $(a_i a)^m \equiv a^m \equiv 1 \bmod p$. Since the polynomial $X^m - 1$ has at most m roots in the finite field $\mathbb{Z}/p\mathbb{Z}$, this is impossible. □

We now give a second proof of the multiplicativity of the numerator of the Legendre symbol using Euler's criterion:

Proposition 3.5. *For integers a, b coprime to p we have*

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

Proof. If $(\frac{a}{p}) = +1$ or $(\frac{b}{p}) = +1$, this actually follows easily from the definitions. If, however, $(\frac{a}{p}) = (\frac{b}{p}) = -1$, then we have to work harder. The following proof covers all cases: we have $(\frac{a}{p}) \equiv a^{(p-1)/2} \bmod p$, $(\frac{b}{p}) \equiv b^{(p-1)/2} \bmod p$, and $(\frac{ab}{p}) \equiv (ab)^{(p-1)/2} \bmod p$. This implies $(\frac{a}{p})(\frac{b}{p}) \equiv (\frac{ab}{p}) \bmod p$, hence p divides the difference $(\frac{a}{p})(\frac{b}{p}) - (\frac{ab}{p})$. But the absolute value of this difference is ≤ 2, and since the only number in this interval that is divisible by p is 0, the difference must be 0. □

Another corollary is the quadratic character of -1:

Proposition 3.6. *For odd primes p we have*

$$\Big(\frac{-1}{p}\Big) = (-1)^{\frac{p-1}{2}}.$$

This is called the first supplementary law of quadratic reciprocity. The proof is easy: by Euler's criterion, we have $(\frac{-1}{p}) \equiv (-1)^{(p-1)/2} \bmod p$. Thus $(\frac{-1}{p}) - (-1)^{(p-1)/2}$ is an integer between -2 and $+2$ that is divisible by p: this implies equality $(\frac{-1}{p}) = (-1)^{(p-1)/2}$.

This simple result allows us to prove that there are infinitely many primes of the form $4n + 1$. We first formulate a little

Lemma 3.7. *If $p > 0$ is an odd prime divisor of an integer of the form $n^2 + 1$, then $p \equiv 1 \bmod 4$.*

Proof. From $p \mid n^2 + 1$ we deduce that $n^2 \equiv -1 \bmod p$. Thus -1 is a quadratic residue modulo p, hence $p \equiv 1 \bmod 4$. □

Corollary 3.8. *There are infinitely many primes of the form $4n + 1$.*

Proof. Assume there are only finitely many primes of the form $4n+1$, say $p_1 = 5, p_2, \ldots, p_n$. Then $N = 4p_1^2 \cdots p_n^2 + 1$ is of the form $4n+1$ and greater than all the primes p_k of this form, hence N must be composite. Now N is odd, hence so is any prime divisor p of N, and since any such p is of the form $4n+1$ by Prop. 3.6, we conclude that $p = p_k$ for some index k. But then $p_k \mid N$ and $p_k \mid N - 1 = 4p_1^2 \cdots p_n^2$, and we get the contradiction that $p_k \mid (N - (N-1)) = 1$. □

Proofs using Primitive Roots

Using the existence of primitive roots modulo p, many of the ad-hoc proofs given above become a lot clearer.

Recall that a primitive root g modulo p has the property that the elements

$$g, g^2, \ldots, g^{p-1} \tag{3.1}$$

represent the $p-1$ elements of $(\mathbb{Z}/p\mathbb{Z})^\times$. Since $g^{p-1} \equiv 1 \bmod p$, the powers g^j for $1 \le j < p-1$ cannot be congruent to 1 modulo p (because in this case the residue class $1 \bmod p$ would be represented twice by the elements in (3.1), hence some other element would be missing.

In particular, $g^{(p-1)/2} \not\equiv 1 \bmod p$. But since $(g^{(p-1)/2})^2 \equiv g^{p-1} \equiv 1 \bmod p$ and since the congruence $x^2 \equiv 1 \bmod p$ has exactly two solutions modulo primes p, namely $x \equiv \pm 1 \bmod p$, we conclude

Proposition 3.9. *If g is a primitive root modulo p, then*

$$g^{\frac{p-1}{2}} \equiv -1 \bmod p.$$

In particular, primitive roots are quadratic nonresidues modulo p.

Note that there are $\frac{p-1}{2}$ quadratic residues, but only $\phi(p-1)$ primitive roots modulo p. Since, in general, $\phi(p-1) < \frac{p-1}{2}$, there must be more quadratic nonresidues than primitive roots.

We can also immediately prove the first supplementary law: assume first that $x^2 \equiv -1 \bmod p$ has a solution. Raising both sides to the power of $\frac{p-1}{2}$ and using Fermat's little theorem we get $1 \equiv (-1)^{(p-1)/2} \bmod p$. Thus $\frac{p-1}{2}$ must be even, and this shows that $p \equiv 1 \bmod 4$. Conversely, assume that $p \equiv 1 \bmod 4$ an write $p = 4n+1$. Let g be a primitive roots modulo p and put $x = g^n$. Then $x^2 = g^{2n} = g^{(p-1)/2} \equiv -1 \bmod p$ by Prop. 3.9.

Next we claim:

Proposition 3.10. *Let g be a primitive root modulo p. Then the quadratic residues modulo p are represented by the even powers $g^2, g^4, \ldots, g^{p-1}$ of g modulo p, and the quadratic nonresidues by the odd powers $g, g^3, \ldots, g^{p-2}$.*

Proof. Since the residue classes $g^2, g^4, \ldots, g^{p-1} \bmod p$ are obviously squares, they are quadratic residues; since g is a primitive root, they are all distinct modulo p. Since there are exactly $\frac{p-1}{2}$ quadratic residues, the remaining powers $g, g^3, \ldots, g^{p-2}$ must represent the nonresidues modulo p. □

Now let us see why Euler's criterium is true: if $(\frac{a}{p}) = +1$, then $a \equiv g^{2k} \bmod p$, hence $a^{(p-1)/2} \equiv (g^{p-1})^k \equiv 1 \bmod p$; if $(\frac{a}{p}) = -1$, then $a \equiv g^{2k+1} \bmod p$, hence $a^{(p-1)/2} \equiv (g^{p-1})^k g^{(p-1)/2} \equiv -1 \bmod p$ because of Prop. 3.9.

The multiplicativity of the Legendre symbol is now a trivial consequence of Euler's criterium: we have $(\frac{a}{p})(\frac{b}{p}) \equiv a^{(p-1)/2} b^{(p-1)/2} = (ab)^{(p-1)/2} \equiv (\frac{ab}{p}) \bmod p$, and now the usual argument (both sides are ± 1, and they are congruent modulo an odd prime, hence they must be equal) shows that $(\frac{a}{p})(\frac{b}{p}) = (\frac{ab}{p})$.

3.2 Gauss's Lemma

Gauss's Lemma is a result that can be used to prove the quadratic reciprocity law we will state in the next section. Although we will prove the reciprocity law with different means, we still will present Gauss's Lemma and use it to prove the supplementary laws.

Recall how we proved Fermat's Little Theorem: we took a complete set of nonzero residue classes $\{1, 2, \ldots, p-1\}$, multiplied everything by a, and pulled out the factor a^{p-1}. For quadratic reciprocity, Euler's criterion suggests that we would like to pull out a factor $a^{(p-1)/2}$. That's what made Gauss introduce a halfsystem modulo p: this is any set $A = \{a_1, \ldots, a_m\}$ of representatives for residue classes modulo $p = 2m + 1$ with the following properties:
a) the a_j are distinct modulo p, that is: if $a_i \equiv a_j \bmod p$, then $i = j$;
b) every integer is either congruent modulo p to a_i or to $-a_i$ for some $1 \leq i \leq \frac{p-1}{2}$.

In other words: a halfsystem A is any set of integers such $A \cup -A$ is a complete set of nonzero residue classes modulo p. A typical halfsystem modulo p is the set $A = \{1, 2, \ldots, \frac{p-1}{2}\}$.

Now consider the prime $p = 13$, choose $A = \{1, 2, 3, 4, 5, 6\}$, and look at $a = 2$. Proceeding as in the proof of Fermat's Little Theorem, we multiply everything in sight by 2 and find

$$\begin{aligned} 2 \cdot 1 &\equiv +2 \bmod 13, \\ 2 \cdot 2 &\equiv +4 \bmod 13, \\ 2 \cdot 3 &\equiv +6 \bmod 13, \\ 2 \cdot 4 &\equiv -5 \bmod 13, \\ 2 \cdot 5 &\equiv -3 \bmod 13, \\ 2 \cdot 6 &\equiv -1 \bmod 13. \end{aligned}$$

Thus three products still lie in A, while three others lie in $-A$. Thus there is an odd number of sign changes, and 2 is a quadratic nonresidue. This is no surprise: multiplying the congruences we find $2^6 \cdot 6! \equiv (-1)^3 \cdot 6! \bmod 13$, hence by Euler's criterium $(\frac{2}{13}) \equiv 2^6 \equiv -1 \bmod 13$.

What about $a = 3$? Here we find

$$\begin{aligned} 3 \cdot 1 &\equiv +3 \bmod 13, \\ 3 \cdot 2 &\equiv +6 \bmod 13, \\ 3 \cdot 3 &\equiv -4 \bmod 13, \\ 3 \cdot 4 &\equiv -1 \bmod 13, \\ 3 \cdot 5 &\equiv +2 \bmod 13, \\ 3 \cdot 6 &\equiv +5 \bmod 13. \end{aligned}$$

Here the number of sign changes is even (there are two), and 3 is a quadratic residue modulo 13.

Gauss realized that this is not an accident:

Lemma 3.11 (Gauss's Lemma)**.** *Let $p = 2n + 1$ be an odd prime, put $A = \{a_1, \ldots, a_n\}$, and let a be an integer not divisible by p. Write*

$$a_i a \equiv (-1)^{s(i)} a_{t(i)} \bmod p \tag{3.2}$$

for every $a_i \in A$, where $s(i) \in \{0, 1\}$ and $t(i) \in \{1, 2, \ldots, n\}$. Then

$$a^n \equiv \prod_{i=1}^{n} (-1)^{s(i)} \bmod p.$$

Thus a is a quadratic residue or nonresidue modulo p according as the number of sign changes is even or odd. The proof is quite simple:

Proof. Observe that the $a_{t(i)}$ in (3.2) run through A if the a_i do, that is: the $a_{t(i)}$ are just the a_i in a different order. In fact, if we had $a_i a \equiv (-1)^{s(i)} a_{t(i)} \bmod p$ and $a_k a \equiv (-1)^{s(k)} a_{t(k)} \bmod p$ with $a_{t(i)} = a_{t(k)}$, then dividing the first congruence by the second gives $a_i / a_k \equiv (-1)^{s(i)-s(k)} \bmod p$, that is, we have $a_i \equiv \pm a_k \bmod p$ for some choice of sign. But this implies $a_i = a_k$ since $1 \leq a_i, a_k \leq \frac{p-1}{2}$.

Now we apply the usual trick: if two sets of integers coincide, then the product over all elements must be the same. In our case, this means that $\prod_{i=1}^{n} a_i a \equiv \prod_{i=1}^{n} (-1)^{s(i)} a_{t(i)} \bmod p$. The left hand side equals $(a_1 a) \cdot (a_2 a) \cdots (a_n a) = a^n \prod_{i=1}^{n} a_i$, whereas the right hand side is $\prod_{i=1}^{n} (-1)^{s(i)} \cdot \prod_{i=1}^{n} a_{t(i)}$. But we have $\prod_{i=1}^{n} a_{t(i)} = \prod_{i=1}^{n} a_i$ by the preceding paragraph. Thus we find $a^n \prod_{i=1}^{n} a_i \equiv \prod_{i=1}^{n} (-1)^{s(i)} \prod_{i=1}^{n} a_i$, and since the product over the a_i is coprime to p, it may be canceled; this proves the claim. □

In the last section we have proved the first supplementary law, which tells us when -1 is a square modulo p. The second supplementary law will tell us when 2 is a square modulo p. Let us first make a short table which will tell us what happens for small primes p:

p	3	5	7	11	13	17	19	23	29	31
$(2/p)$	-1	-1	$+1$	-1	-1	$+1$	-1	$+1$	-1	$+1$
$\sqrt{2}$	$-$	$-$	± 3	$-$	$-$	± 6	$-$	± 5	$-$	± 8

Thus 2 is a quadratic residue modulo 7, 17, 23, and 31; among the primes in this table, these are exactly the primes of the form $p \equiv \pm 1 \bmod 8$. Thus we conjecture:

Proposition 3.12. *The prime* 2 *is a quadratic residue modulo an odd prime* p *if and only if* $p \equiv \pm 1 \bmod 8$. *In other words: we have* $(\frac{2}{p}) = (-1)^{(p^2-1)/8}$.

The fact that the second claim is equivalent to the first is easy to check: Basically, the proof boils down to the following table:

$a \bmod 8$	1	3	5	7
$\frac{1}{8}(a^2-1) \bmod 2$	0	1	1	0

Let us now use Gauss's Lemma to give a proof for Prop. 3.12. We have to count the number of sign changes when we multiply the "half system" $A = \{1, 2, \ldots, \frac{p-1}{2}\}$ by 2.

1. Assume first that $p = 4k+1$, i.e. $\frac{p-1}{2} = 2k$.

$$\begin{aligned}
[1] \cdot [2] &= [2] \\
[2] \cdot [2] &= [4] \\
\ldots &= \ldots \\
[k] \cdot [2] &= [2k] \\
[k+1] \cdot [2] &= [2k+2] \;=\; -[2k-1] \\
\ldots &= \ldots \\
[2k] \cdot [2] &= [4k] \;=\; -[1]
\end{aligned}$$

Here $2a \le 2k$ for $a < k$, that is for $a = 1, 2, \ldots, k$, so there are no sign changes at all for these a. If $k < a \le 2k$, however, then $2k < 2a \le p-1$, hence $1 \le p - 2a < p - 2k = 2k+1$, which implies that there are sign changes for each a in this interval. Since there are exactly k such a, Gauss's Lemma says that $(\frac{2}{p}) = (-1)^k$; we only have to check that $k \equiv \frac{p^2-1}{8} \bmod 2$. But this follows from $\frac{p^2-1}{8} = \frac{1}{8}(p-1)(p+1) = \frac{1}{8} \cdot 4k(4k+2) = k(2k+1)$.

2. Now assume that $p = 4k-1$; then there are no sign changes whenever $1 \le a \le k-1$, and there are exactly k sign changes for $k \le a < 2k$, so again we have $(\frac{2}{p}) = (-1)^k$. But now $\frac{p^2-1}{8} = \frac{1}{8}(p-1)(p+1) = (2k-1)k$ shows that $k \equiv \frac{p^2-1}{8} \bmod 2$.

The second supplementary law can be used to prove that there are infinitely many primes of the form $p \equiv \pm 1 \bmod 8$.

For a different application, consider the Mersenne numbers M_q, where q is odd and $p = 2q+1$ is prime. If $q \equiv 3 \bmod 4$, then $p \equiv 7 \bmod 8$, hence $(2/p) = 1$.

By Euler's criterion, this means that $2^q = 2^{(p-1)/2} \equiv 1 \bmod p$, and this in turn shows that $p \mid M_q$.

Corollary 3.13. *If $p = 2q + 1 \equiv 7 \bmod 8$ is prime, then $p \mid M_q$, the q-th Mersenne number.*

In particular, $23 \mid M_{11}$ and $83 \mid M_{41}$. Thus some Mersenne numbers can easily be seen to be composite. There are similar (but more complicated) rules for $p \mid M_q$ when $p = 4q + 1$; in this case, we have to study $2^{(p-1)/4} \bmod p$, which leads us to quartic reciprocity. There is a quartic reciprocity law, but this cannot be formulated in $\mathbb{Z}$: Gauss showed in 1832 that one has to enlarge $\mathbb{Z}$ to the ring $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}, i^2 = -1\}$ for doing that.

3.3 The Quadratic Reciprocity Law

The quadratic reciprocity law connects the Legendre symbols (p/q) and (q/p) for odd primes. Here it is:

Theorem 3.14. *Let p and q be distinct odd primes. Then*

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

For the proof, we need a few things. Let us start with

Proposition 3.15 (Wilson's Theorem)**.** *For primes p, we have $(p-1)! \equiv -1 \bmod p$.*

Proof. Let p be a prime; the claim is trivial if $p = 2$, so assume that p is odd. The idea is to look at pairs of the elements of $(\mathbb{Z}/p\mathbb{Z})^\times$. In fact, for every $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ there is an element $a^{-1} \in (\mathbb{Z}/p\mathbb{Z})^\times$ such that $a \cdot a^{-1} \equiv 1 \bmod p$. In general, $[a]$ and $[a^{-1}]$ are different: $[a] = [a^{-1}]$ implies $[a^2] = [1]$, so this can only happen (and does in fact happen) if $[a] = [1]$ or $[a] = [-1] = [p-1]$ (here we use that $\mathbb{Z}/p\mathbb{Z}$ is a field; in fields, polynomials of degree 2 such as $x^2 - 1$ have at most 2 roots).

Thus $(\mathbb{Z}/p\mathbb{Z})^\times \setminus \{[-1], [+1]\}$ is the union of pairs $\{[a], [a^{-1}]\}$ with $[a] \neq [a^{-1}]$, hence the product over all elements of $(\mathbb{Z}/p\mathbb{Z})^\times \setminus \{[-1], [+1]\}$ must be $[1]$. We can get $[(p-1)!]$ by multiplying this product with the two missing classes $[1]$ and $[-1]$, and this gives the claimed result $[(p-1)!] = [-1]$. □

There is another proof of Wilson's theorem that uses Fermat's Little Theorem:

Second Proof of Wilson's Theorem. Consider the polynomial $f(X) = X^{p-1} - 1$ over $\mathbb{F}_p[X]$ (that is, we think of the coefficients as residue classes modulo p. Since f has the roots $X = 1, 2, \ldots, p-1$ by Fermat's little theorem, and since we are working over the *field* $\mathbb{F}_p$, we must have

$$X^{p-1} - 1 \equiv (X-1)(X-2)\cdots(X-(p-1)) \bmod p.$$

Comparing the constant terms yields Wilson's Theorem. □

Now let us see how to prove the reciprocity law. Since we are comparing the Legendre symbols $(\frac{p}{q})$ and $(\frac{q}{p})$, which tell us something about the rings $\mathbb{Z}/q\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z}$, respectively, it seems only natural to look for a ring that "contains both", namely $\mathbb{Z}/pq\mathbb{Z}$. The simplest proofs of the quadratic reciprocity law are of this type, and this includes the following proof by Kim (2004).

He sets $A = \{a : 1 \leq a \leq \frac{pq-1}{2}, \gcd(a, pq) = 1\}$ and computes $\prod_{a\in A} a$ modulo p, q, and pq.

Lemma 3.16. *For all distinct odd primes p and q we have*

$$\prod_{a\in A} \equiv \left(\frac{-1}{q}\right)\left(\frac{q}{p}\right) \bmod p.$$

Proof. Consider the set $S = \{a : 1 \leq a \leq \frac{pq-1}{2}, \gcd(a, p) = 1\}$. Clearly $A \subset S$, and $T = S \setminus A$ consists of all elements of S that are multiples of q. Thus $T = \{q, 2q, \ldots, \frac{p-1}{2}q\}$.

We now compute $\prod_{a\in S} a \bmod p$. To this end, we observe that the elements of S are

$$\begin{array}{cccc}
1 & 2 & \ldots & p-1 \\
p+1 & p+2 & \ldots & 2p-1 \\
2p+1 & 2p+2 & \ldots & 3p-1 \\
 & & \vdots & \\
\frac{q-3}{2}p+1 & \frac{q-3}{2}p+2 & \ldots & \frac{q-3}{2}p+p-1 \quad = \frac{q-1}{2}p-1 \\
\frac{q-1}{2}p+1 & \frac{q-1}{2}p+2 & \ldots & \frac{q-1}{2}p+\frac{p-1}{2}
\end{array}$$

The first $\frac{q-1}{2}$ rows all reduce to 1, 2, …, $p-1$ modulo p, the last row to 1, 2, …, $\frac{p-1}{2}$ mod p. Thus

$$\prod_{a\in S} a \equiv [(p-1)!]^{\frac{q-1}{2}} \Big(\frac{p-1}{2}\Big)! \equiv (-1)^{\frac{q-1}{2}} \Big(\frac{p-1}{2}\Big)! \bmod p.$$

Now

$$\prod_{a\in T} a = q \cdot 2q \cdots \frac{p-1}{2}q = q^{\frac{p-1}{2}} \Big(\frac{p-1}{2}\Big)! \equiv \Big(\frac{q}{p}\Big)\Big(\frac{p-1}{2}\Big)! \bmod p,$$

and thus

$$\prod_{a\in A} a = \frac{\prod_{a\in S} a}{\prod_{a\in T} a} \equiv \frac{(-1)^{\frac{q-1}{2}}(\frac{p-1}{2})!}{(\frac{q}{p})(\frac{p-1}{2})!} \equiv \Big(\frac{-1}{q}\Big)\Big(\frac{q}{p}\Big) \bmod p$$

as claimed. □

Note that, by switching the roles of p and q, we also have

$$\prod_{a\in A} \equiv \left(\frac{-1}{p}\right)\left(\frac{p}{q}\right) \bmod q.$$

Next we would like to derive some information about $\prod_{a\in A} a$ modulo pq. To this end, we will need

Lemma 3.17. *There are exactly four solutions of* $x^2 \equiv 1 \bmod pq$.

Proof. Recall that $x^2 \equiv 1 \bmod p$ has exactly two solutions, namely $x \equiv \pm 1 \bmod p$, and this is proved by observing that $p \mid (x^2-1) = (x-1)(x+1)$ implies $p \mid (x-1)$ or $p \mid (x+1)$.

Now $x^2 \equiv 1 \bmod pq$ is equivalent to $pq \mid (x-1)(x+1)$, and now there are four possibilities:

1. $pq \mid (x-1)$: then $x \equiv 1 \bmod pq$.
2. $pq \mid (x+1)$: then $x \equiv -1 \bmod p$.
3. $p \mid (x-1)$, $q \mid (x+1)$: this determines a unique residue class modulo pq. In fact, use Bezout to write $pm + qn = 1$; then $x \equiv qn - pm \bmod pq$.
4. $q \mid (x-1)$, $p \mid (x+1)$: then $x \equiv pm - qn \bmod pq$.

This concludes the proof. □

Since we have proved that $x = \prod_{a\in A} a$ satisfies $x \equiv \pm 1 \bmod p$ and $x \equiv \pm 1 \bmod q$, there are four possibilities for $x \bmod pq$. Now we claim

Lemma 3.18. *We have* $\prod_{a\in A} a \equiv \pm 1 \bmod pq$ *if and only if* $p \equiv q \equiv 1 \bmod 4$.

Proof. We first observe that, by the Chinese remainder theorem, the congruence $x^2 \equiv -1 \bmod pq$ has a solution if and only if there are solutions mod p and mod q, that is, if and only if $p \equiv q \equiv 1 \bmod 4$. In this case, there are exactly four solutions.

For each $a \in A$ consider the inverse $a' \equiv a^{-1} \bmod pq$. Since $\gcd(a', pq) = 1$, we have $a \in A$ or $-a' \in A$, in other words: for every $a \in A$ there is an $a' \in A$ such that $a'a \equiv \pm 1 \bmod pq$.

Now $a' = a$ if and only if $a^2 \equiv \pm 1 \bmod pq$. There are four solutions $\pm 1, \pm N$ of $a^2 \equiv 1 \bmod pq$, among which only 1 and N, say, are in A. Unless $p \equiv q \equiv 1 \bmod 4$, these are the only elements for which $a' = a$. Thus we find $\prod_{a\in A} \equiv \pm N \bmod pq$.

If $p \equiv q \equiv 1 \bmod p$, then there are four solutions I, $-I$, NI, $-NI$ of $x^2 \equiv -1 \bmod p$, among which we can choose I and NI (or I and $-NI$) in A. Thus $\prod_{a\in A} \equiv \pm N \cot I \cdot NI \equiv \pm N^2 I^2 equiv \pm 1 \bmod pq$. □

Now we are ready for the

Proof of the Quadratic Reciprocity Law. We know

$$\prod_{a\in A} a \equiv \begin{cases} (\frac{-1}{q})(\frac{q}{p}) \bmod p, \\ (\frac{-1}{p})(\frac{p}{q}) \bmod q. \end{cases}$$

Lemma 3.18 tells us that the expressions on the right hand side are equal if $p \equiv q \equiv 1 \bmod 4$, and that they differ by a minus sign otherwise. Since $(-1)^{\frac{p+1}{2}\frac{q+1}{2}} = -1$ if and only if $p \equiv q \equiv 1 \bmod 4$, this means

$$\Big(\frac{p}{q}\Big)\Big(\frac{p}{q}\Big) = -(-1)^{\frac{p+1}{2}\frac{q+1}{2}}\Big(\frac{-1}{p}\Big)\Big(\frac{-1}{q}\Big) = -(-1)^{\frac{p+1}{2}\frac{q+1}{2}}(-1)^{\frac{p-1}{2}}(-1)^{\frac{q-1}{2}}.$$

Now

$$\begin{aligned} 1 + \frac{p+1}{2}\frac{q+1}{2} + \frac{p-1}{2} + \frac{q-1}{2} &\equiv \frac{p+1}{2}\frac{q+1}{2} - \Big(\frac{p-1}{2} + \frac{q-1}{2} + 1\Big) \\ &= \frac{p+1}{2}\frac{q+1}{2} - \Big(\frac{p-1}{2} + \frac{q-1}{2} + 1\Big) \\ &= \frac{pq-p-q+1}{4} = \frac{p-1}{2}\frac{q-1}{2} \bmod 2, \end{aligned}$$

and this finally proves the reciprocity law. □

Some simple Consequences

Proposition 3.19. *We have*

$$\Big(\frac{-3}{p}\Big) = \begin{cases} +1 & \text{if } p \equiv 1 \bmod 3, \\ -1 & \text{if } p \equiv 2 \bmod 3. \end{cases}$$

Proof. We find $(\frac{-3}{p}) = (\frac{-1}{p})(\frac{3}{p}) = (-1)^{(p-1)/2}(\frac{p}{3})(-1)^{\frac{p-1}{2}\frac{3-1}{2}} = (\frac{p}{3})$. But $(\frac{p}{3}) = (\frac{1}{3}) = 1$ if $p \equiv 1 \bmod 3$, and $(\frac{p}{3}) = (\frac{2}{3}) = -1$ if $p \equiv 2 \bmod 3$. □

In a similar way we can prove

Proposition 3.20. *We have*

$$\Big(\frac{5}{p}\Big) = \begin{cases} +1 & \text{if } p \equiv \pm 1 \bmod 5, \\ -1 & \text{if } p \equiv \pm 2 \bmod 5. \end{cases}$$

Other applications include the simple calculation of Legendre symbols: for example, we find $(\frac{11}{17}) = (\frac{17}{11}) = (\frac{6}{11}) = (\frac{2}{11})(\frac{3}{11}) = -(\frac{3}{11}) = (\frac{11}{3}) = (\frac{2}{3}) = -1$. Thus 11 is a quadratic nonresidue modulo 17.

3.4 The Jacobi Symbol

The Legendre symbol $(\frac{a}{p})$ can be generalized to composite values of p: if $b = p_1 \cdots p_r$ is a product of odd primes, then we put

$$\Big(\frac{a}{b}\Big) = \Big(\frac{a}{p_1}\Big) \cdots \Big(\frac{a}{p_r}\Big).$$

Thus $(\frac{2}{15}) = (\frac{2}{3})(\frac{2}{5}) = (-1)(-1) = +1$. Note, however, that 2 is not a quadratic residue modulo 15. In fact, we only have

Proposition 3.21. *If $(\frac{a}{b}) = -1$, then a is a quadratic nonresidue modulo b.*

Proof. If $(\frac{a}{b}) = -1$ and $b = \prod p$, then $\prod(\frac{a}{p}) = -1$, and this implies that $(fracap) = -1$ for at least one prime dividing b. Now $a \equiv x^2 \bmod b$ implies $a \equiv x^2 \bmod p$, hence a is a quadratic nonresidue modulo b. □

We also can generalize the first supplementary law:

Proposition 3.22. *We have*

$$\Big(\frac{-1}{a}\Big) = (-1)^{\frac{a-1}{2}}$$

for all odd integers $a > 0$.

Proof. Write $n = p_1 \cdots p_r$; then

$$\Big(\frac{-1}{n}\Big) = \Big(\frac{-1}{p_1}\Big) \cdots \Big(\frac{-1}{p_r}\Big) = (-1)^{\frac{p_1-1}{2}+\ldots+\frac{p_r-1}{2}}.$$

Thus it remains to show that

$$\frac{n-1}{2} \equiv \frac{p_1-1}{2} + \ldots + \frac{p_r-1}{2} \bmod 2. \tag{3.3}$$

This is done by induction. We start with the observation that $(a-1)(b-1) \equiv 0 \bmod 4$ for odd integers a, b, hence $ab-1 \equiv (a-1)+(b-1) \bmod 4$, and dividing by 2 gives

$$\frac{ab-1}{2} \equiv \frac{a-1}{2} + \frac{b-1}{2} \bmod 2.$$

Now use induction. □

Now let us treat the reciprocity law similarly.

Theorem 3.23 (Reciprocity Law for Jacobi Symbols)**.** *If m and n are coprime positive odd integers, then*

$$\Big(\frac{m}{n}\Big)\Big(\frac{n}{m}\Big) \;=\; (-1)^{\frac{m-1}{2}\,\frac{n-1}{2}}.$$

Moreover, we have the supplementary laws

$$\Big(\frac{-1}{n}\Big) \;=\; (-1)^{\frac{n-1}{2}}, \quad \Big(\frac{2}{n}\Big) \;=\; (-1)^{\frac{n^2-1}{8}}.$$

Proof. Write $m = p_1 \cdots p_r$ and $n = q_1 \cdots q_s$; then

$$\Big(\frac{m}{n}\Big)\Big(\frac{n}{m}\Big) = \prod_{i=1}^{r}\prod_{j=1}^{s}\Big(\frac{p_i}{q_j}\Big)\Big(\frac{q_j}{p_i}\Big) = \prod_{i=1}^{r}\prod_{j=1}^{s}(-1)^{(p_i-1)(q_j-1)/4},$$

and our claim will follow if we can prove that

$$\frac{m-1}{2}\frac{n-1}{2} \equiv \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{p_i-1}{2}\frac{q_j-1}{2} \bmod 4.$$

But this follows by multiplying the two congruences you get by applying (3.3) to m and n.

Finally, consider the second supplementary law. Similar to the above, everything boils down to showing

$$\frac{n^2-1}{8} \equiv \frac{p_1^2-1}{8} + \ldots + \frac{p_r^2-1}{8} \bmod 2.$$

Now clearly $16 \mid (a^2-1)(b^2-1)$ (as a matter of fact, even this product is even divisible by 64), hence

$$(ab)^2 - 1 \equiv a^2 - 1 + b^2 - 1 \bmod 16.$$

Now induction does the rest. □

Exercises

3.1 Use Gauss's Lemma to prove that $(\frac{-2}{p}) = +1$ or -1 according as $p \equiv 1, 3 \bmod 8$ or $p \equiv 5, 7 \bmod 8$.

3.2 Show that every prime $p \neq 3$ dividing a number of the form $4n^2 + 3$ satisfies $p \equiv 1 \bmod 3$.

3.3 Show that there are infinitely many primes $p \equiv 1 \bmod 3$.

3.4 All primes dividing a number of the form $n^2 + 1$ are congruent to 1 mod 4.

3.5 Show that all odd prime divisors of $999998 = 1000^2 - 2$ satisfy $p \equiv \pm 1 \bmod 8$

3.6 Show that $y^2 = x^3 + 7$ has no integer solutions.

Hints: (This proof is due to V.A. Lebesgue)

1. Show that x is odd.
2. Write the equation as $y^2 + 1 = x^3 + 8$ and factor the right hand side.
3. Show that the quadratic factor is divisible by some prime $p \equiv 3 \bmod 4$
4. Look at the left hand side.

3.7 Generalize the preceding exercise to an infinite family of diophantine equations $y^2 = x^3 + c$.

3.8 (This is a conjecture by Euler) Prove that if $p \equiv 1 \bmod 4$ is prime and $a = \frac{p-1}{4} - n - n^2$, then $(q/p) = +1$ for every $q \mid a$.

3.9 (Euler) If $p \equiv 1 \bmod 4$ is prime, then $\frac{p-1}{4} - n(n+1)$ is a quadratic residue modulo p for every integer n.

3.10 (Euler) If $q \equiv 3 \bmod 4$ is prime, then $\frac{q+1}{4} + n(n+1)$ is a quadratic residue modulo q for every integer n.

3.11 Show that

$$\left(\frac{3}{p}\right) = \begin{cases} +1 & \text{if } p \equiv \pm 1 \bmod 12, \\ -1 & \text{if } p \equiv \pm 5 \bmod 12. \end{cases}$$

3.12 Show that $\sum_{a=1}^{p-1} (\frac{a}{p}) = 0$.

3.13 Let $p = a^2 + 4b^2$ be a prime. Show that $(\frac{a}{p}) = +1$.

3.14 (Bork) If q and $p = q + 4$ are prime, then $(p/q) = 1$.

3.15 (Bickmore) Since $S_p = 2^{2p} + 1$ is the sum of two squares, so is each of its factors. Verify that, for $p = 2m + 1$, $S_p = A_p B_p$ for $A_p = 2^p - 2^{m+1} + 1$ and $B_p = 2^p + 2^{m+1} + 1$, and write A_p and B_p as a sum of two squares. Use the quadratic reciprocity law to prove that $5 \mid A_p \iff (2/p) = -1$ and $5 \mid B_p \iff (2/p) = +1$.

3.16 Show that $2^{340} \equiv 1 \bmod 341$ (Hint: $341 \cdot 3 = 1023$). Also show that 341 is not prime.

3.17 Show that, for Fermat primes $p = F_n = 2^{2^n} + 1$, we have $\phi(p-1) = \frac{p-1}{2}$. Conclude that every quadratic nonresidue modulo p is a primitive root mod p.

3.18 Show that 3 is a primitive root modulo p for every for every Fermat prime F_n with $n > 0$.

3.19 Let $n = 4m^2 + 3$, where m is an integer not divisible by 3. Show that there exists a prime $p \mid n$ with $p \equiv 7 \bmod 12$.

3.20 Show that there are infinitely many primes $p \equiv 7 \bmod 12$.

3.21 Show that all prime factors p of $n = m^4 - m^2 + 1$ satisfy $p \equiv 1 \bmod 4$. Show that $p \equiv 1 \bmod 12$ whenever $3 \mid m$.

3.22 Show that there are infinitely many primes $p \equiv 1 \bmod 12$.

3.23 Prove that if p and $q = 2p + 1$ are both odd primes, then $g = -4$ is a primitive root mod q.

3.24 Prove that if $p \equiv 1 \bmod 4$ is a prime, then -4 and $\frac{p-1}{4}$ are both quadratic residues of p.

3.25 Assume that p and $q = 4p + 1$ are both primes; show that any quadratic non-residue of q is either a primitive root mod q or has order 44 mod q. Also show that 2 is a primitive root mod q.

3.26 The theory of biquadratic residues is much more complicated than the theory of quadratic residues. This exercise gives the analog of the second supplementary law: the congruence $x^4 \equiv 2 \bmod p$ is solvable for a prime $p \equiv 1 \bmod 8$ if and only if $p = x^2 + 64y^2$.

1. Prove that $x^4 \equiv 2 \bmod p$ is solvable for a prime $p \equiv 1 \bmod 8$ if and only if $2^{(p-1)/4} \equiv 1 \bmod p$.
2. By Fermat's two-squares-theorem we can write $p = a^2 + b^2$ with a odd and b even. We have already proved that $(\frac{a}{p}) = +1$. Show that $(\frac{a+b}{p}) = (\frac{2}{a+b})$.
3. Write $a \equiv bf \bmod p$ for some integer f. Show that $f^2 \equiv -1 \bmod p$.
4. Show that $2^{(p-1)/4} \equiv f^{ab/2} \bmod p$. Hint: $(1 + f)^2 \equiv 2f \bmod p$.
5. Prove Gauss's claim. (This is Dirichlet's proof.)

Chapter 4

Binary Quadratic Forms

A binary quadratic form is an expression

$$Q(X,Y) = AX^2 + BXY + CY^2,$$

where A, B, C are integers; we will often abbreviate this as $Q = (A, B, C)$. The integer $\Delta = B^2 - 4AC$ is called the discriminant of the form Q. For example, $(1, 0, 1)$ is the form $X^2 + Y^2$ with discriminant $\Delta = -4$. We say that an integer n is represented by Q if there exist integers x, y such that $Q(x, y) = n$. The question we will study is: which integers (and, more specifically, which primes) are represented by a given form Q?

For the simplest form $X^2 + Y^2$, this question was answered by Fermat: a prime p is represented by $Q = (1, 0, 1)$ if and only if $p = 2$ or $p \equiv 1 \bmod 4$. In particular, all the primes p with $(\frac{\Delta}{p}) = +1$ are represented by Q.

In this chapter, we will develop a method that allows us to answer similar questions for a large class of quadratic forms.

4.1 The Action of $\mathrm{SL}_2(\mathbb{Z})$ on Forms

Consider a quadratic form $Q = (A, B, C)$, and let r, s, t, u be integers. Putting $X = rZ + sW$ and $Y = tZ + uW$ we get

$$\begin{aligned} Q(X,Y) &= Q(rZ + sW, tZ + uW) \\ &= A(rZ + sW)^2 + B(rZ + sW)(tZ + uW) + C(tZ + uW)^2 \\ &= A'Z^2 + B'ZW + C'W^2, \end{aligned}$$

where A', B', C' are integers defined by

$$\begin{aligned} A' &= Ar^2 + Brt + Ct^2, \\ B' &= 2(Ars + Ctu) + B(ru + st), \\ C' &= As^2 + Bsu + Cu^2. \end{aligned}$$

Thus $Q'(Z,W) = Q(rZ+sW, tZ+uW)$.

How are the forms Q and $Q' = (A', B', C')$ related? Assume that $n = Q'(z,w)$ for integers z, w; then $Q'(z,w) = Q(x,y)$ for

$$x = rz + sw, \quad y = tz + uw. \tag{4.1}$$

Thus if n is represented by Q', then n is also represented by Q.

What about the converse? The converse will hold if we can write z and w as $\mathbb{Z}$-linear combinations of x and y. Solving the linear system (4.1) for z and w we find, using Cramer's rule,

$$z = \frac{\left|\begin{smallmatrix} x & s \\ y & u \end{smallmatrix}\right|}{\left|\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right|}, \quad w = \frac{\left|\begin{smallmatrix} r & x \\ t & y \end{smallmatrix}\right|}{\left|\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right|}.$$

Thus z and w will definitely be integers if $ru - st = \pm 1$, and we have shown

Proposition 4.1. *Let Q be a binary quadratic form, let r, s, t, u be integers with $ru - st = \pm 1$, and define the quadratic form Q' by $Q'(X,Y) = Q(rX+sY, tX+uY)$. Then Q and Q' represent exactly the same integers.*

If Q and Q' satisfy $Q'(X,Y) = Q(rX+sY, tX+uY)$, we will write $Q' = Q|_S$ for the matrix $S = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right)$. For reasons that will become clear only later on, we will now restrict to substitutions satisfying $ru - st = +1$, and introduce the special linear group

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \begin{pmatrix} r & s \\ t & u \end{pmatrix} \;\middle|\; ru - st = +1 \right\}.$$

This is a group because

- it is closed under multiplication: if $S, T \in \mathrm{SL}_2(\mathbb{Z})$, then the product ST also has integral entries and $\det ST = (\det S)(\det T) = +1$;
- there is a unit element, namely the identity matrix $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$;
- every element $S = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ has an inverse, namely $S^{-1} = \left(\begin{smallmatrix} u & -s \\ -t & r \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$;
- multiplication is associative: $R(ST) = (RS)T$ for all $R, S, T \in \mathrm{SL}_2(\mathbb{Z})$.

Example. Consider the form $Q(x,y) = x^2 + y^2$ of discriminant $\Delta = -4$, and the matrix $S = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Then $Q|_S(x,y) = (2x+y)^2 + (x+y)^2 = 5x^2 + 6xy + 2y^2$.

We can bring in some more linear algebra in the following way. To every binary quadratic form (A,B,C) we associate the matrix $M = \left(\begin{smallmatrix} A & B/2 \\ B/2 & C \end{smallmatrix}\right)$ (the occurrence of half-integers made Gauss look only at binary quadratic forms whose middle coefficient B is even); then a simple calculation shows that $Ax^2 + Bxy + Cy^2 = (x,y)M\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. We also see that $\operatorname{disc} Q = B^2 - 4AC = -4 \det M$.

If Q corresponds to M, then $Q|_S$ corresponds to S^tMS, where S^t denotes the transpose of S. This is a simple calculation:

$$\begin{aligned} MS &= \begin{pmatrix} A & \frac{B}{2} \\ \frac{B}{2} & C \end{pmatrix} \begin{pmatrix} r & s \\ t & u \end{pmatrix} = \begin{pmatrix} Ar + \frac{B}{2}t & As + \frac{B}{2}u \\ \frac{B}{2}r + Ct & \frac{B}{2}s + Cu \end{pmatrix} \\ S^tMS &= \begin{pmatrix} r & t \\ s & u \end{pmatrix} \begin{pmatrix} Ar + \frac{B}{2}t & As + \frac{B}{2}u \\ \frac{B}{2}r + Ct & \frac{B}{2}s + Cu \end{pmatrix} \\ &= \begin{pmatrix} Ar^2 + Brt + Ct^2 & Ars + Ctu + \frac{B}{2}(ru + st) \\ Ars + Ctu + \frac{B}{2}(ru + st) & As^2 + Bsu + Cu^2 \end{pmatrix} \end{aligned}$$

From the fact that $Q|_S$ corresponds to S^tMS we deduce that

$$\operatorname{disc} Q|_S = -4 \det S^tMS = -4 \det M (\det S)^2 = \operatorname{disc} Q.$$

Moreover we see that $Q|_{ST}$ corresponds to $(ST)^tM(ST) = T^tS^tPST$, hence $Q|_{ST} = (Q|_S)|_T$: this is usually expressed by saying that $\mathrm{SL}_2(\mathbb{Z})$ acts on quadratic forms from the right.

Using the linear algebra approach, we can also give a new proof for Prop. 4.1. In fact, assume that $n = Q(x, y)$ for integers x, y. If M is the associated matrix, then $n = (x, y)M\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Since $Q|_S$ is associated to S^tMS, we have $n = Q|_S(u, v) = (u, v)S^tMS\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ for the vector $\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) = S^{-1}\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Since $S \in \mathrm{SL}_2(\mathbb{Z})$, we have $S^{-1} \in \mathrm{SL}_2(\mathbb{Z})$ as well, and this means that u and v are integers.

Example. $Q(x, y) = x^2 + y^2$ represents $5 = 2^2 + 1^2$. With $S = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ we have found $Q|_S(x, y) = 5x^2 + 6xy + 2y^2$. Now $S^{-1} = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 2 \end{smallmatrix}\right)$, hence $S^{-1}\left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, and indeed we have $Q|_S(1, 0) = 5$.

We now call two binary quadratic forms Q and Q' *equivalent* (and write $Q' \sim Q$) if there exists a matrix $S \in \mathrm{SL}_2(\mathbb{Z})$ such that $Q' = Q|_S$. This is an equivalence relation because it is

- reflexive: $Q \sim Q$. This follows from $Q = Q|_I$, where $I \in \mathrm{SL}_2(\mathbb{Z})$ is the identity matrix.
- symmetric: $Q \sim Q'$ implies $Q' \sim Q$. In fact, if $Q' = Q|_S$, then $Q = Q'|_T$ for $T = S^{-1}$. This can be seen as follows: if M and M' are the matrices attached to Q and Q', then $M' = S^tMS$, hence $M = (S^{-1})^tM'S^{-1}$.
- transitive: $Q \sim Q'$ and $Q' \sim Q''$ imply $Q \sim Q''$. In fact, we have $M' = S^tMS$ and $M'' = T^tM'T$, hence $M'' = T^tM'T = T^tS^tMST = (ST)^tMST$.

For example we have

- $(A, B, C) \sim (C, -B, A)$ (take $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$).
- $(A, -A, C) \sim (A, A, C)$ (take $S = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$).
- $(A, -B, A) \sim (A, B, A)$ (this is a special case of the first example).

Using the notion of equivalence, we can state the results proved above as follows:

- Equivalent forms represent the same integers.
- Equivalent forms have the same discriminant.

In the next section we will investigate the following problem: given a quadratic form, find an equivalent form with coefficients that are as small as possible.

4.2 Reduction

The reduction theory of binary quadratic forms with negative discriminant differs considerably from that of positive discriminant. We only have time to cover the simpler case of negative discriminants. Thus in this section we will assume that $Q = (A, B, C)$ has $A > 0$ and $\Delta = B^2 - 4AC < 0$. Such forms are positive definite since $4AQ(x, y) = (2AX + BY)^2 - \Delta Y^2$. Moreover, we will only consider primitive forms, that is, forms with $\gcd(A, B, C) = 1$.

We say that a positive definite primitive form $Q = (A, B, C)$ is *reduced* if A, B, C satisfy the following conditions: $|B| \leq A \leq C$, with $B > 0$ if one of the inequalitites is not strict.

It is easy to find a bound for the coefficient A of a reduced form:

Lemma 4.2. *If $Q = (A, B, C)$ is a reduced binary quadratic form with negative discriminant Δ, then $|A| \leq \sqrt{-\Delta/3}$.*

Proof. We know $B^2 \leq A^2$ and $A \leq C$, hence $-\Delta = 4AC - B^2 \geq 4A^2 - A^2 = 3A^2$. □

Corollary 4.3. *There are only finitely many reduced forms of discriminant $\Delta < 0$.*

Proof. There are only finitely many A by Lemma 4.2, hence only finitely many B with $|B| \leq A$. Finally, for each pair (A, B) there is at most one C because $\Delta = B^2 - 4AC$ is fixed. □

Actually, the finiteness is an obvious corollary of the following

Lemma 4.4. *If (A, B, C) is a reduced form of discriminant $\Delta < 0$, then $|B| \leq A \leq \sqrt{-\Delta/3}$ and $C \leq \frac{1-\Delta}{4}$.*

Note that these inequalities are sharp e.g. for the form $(1, 1, 1)$ of discriminant $\Delta = -3$.

Proof. Only the last inequality remains to be proved. From $4AC = B^2 - \Delta$ and the fact that $A > 0$ we get

$$C = \frac{B^2}{4A} - \frac{\Delta}{4A} \leq \frac{A^2}{4A} - \frac{\Delta}{4A} = \frac{A}{4} - \frac{\Delta}{4A}.$$

As a function of A (assuming Δ to be constant), the expression on the right hand side is decreasing in the interval $[1, \sqrt{-\Delta}\,]$, hence attains its maximum at the boundary $A = 1$. This implies the claim. □

The number of reduced forms of given discriminant $\Delta < 0$ is called the *class number* $h(\Delta)$.

Since discriminants satisfy $\Delta = B^2 - 4AC \equiv B^2 \equiv 0, 1 \bmod 4$, every discriminant has the form $\Delta = -4m$ or $\Delta = 1 - 4m$. The forms $Q_0 = (1, 0, m)$ of discriminant $\Delta = -4m$ and $Q_0 = (1, 1, m)$ of discriminant $\Delta = 1 - 4m$ are reduced; they are called the principal form of discriminant Δ. We have $h(\Delta) = 1$ if and only if Q_0 is the only reduced form of discriminant Δ.

As an example, let us compute the class number $h(-20)$. We know that $0 < A < \sqrt{20/3} < 3$, hence $A \in \{1, 2\}$. Moreover, $-20 = B^2 - 4AC$ shows that B must be even. The following table then lists all possibilities:

A	B	forms
1	0	$x^2 + 5y^2$
2	0	$- - -$
	2	$2x^2 + 2xy + 3y^2$

Thus there are only two reduced forms, and $h(-20) = 2$.

It is quite easy to compute all reduced forms of small discriminant:

Δ	$h(\Delta)$	reduced forms
-3	1	$x^2 + xy + y^2$
-4	1	$x^2 + y^2$
-7	1	$x^2 + xy + 2y^2$
-8	1	$x^2 + 2y^2$
-11	1	$x^2 + xy + 3y^2$
-12	1	$x^2 + 3y^2$
-15	2	$x^2 + xy + 4y^2, 2x^2 + xy + 2y^2$
-16	1	$x^2 + 4y^2$
-19	1	$x^2 + xy + 5y^2$
-20	2	$x^2 + 5y^2, 2x^2 + 2xy + 3y^2$
-23	3	$x^2 + xy + 6y^2, 2x^2 \pm xy + 3y^2$
-24	2	$x^2 + 6x^2, 2x^2 + 3y^2$
-27	1	$x^2 + xy + 7y^2$

Here is a more complex example: let us explicitly compute the class number for $\Delta = -4 \cdot 65$. We know that $|A| \le \sqrt{-\Delta/3} < 10$. Thus we have $-9 \le A < B \le 9 \le C$ and $-\Delta = 260 = 4AC - B^2$. Clearly $B = 2b$ is even, and we have $65 = AC - b^2$. Now we go through the individual cases; the congruence $65 \equiv b^2 \bmod A$ will occasionally help us to save work.

- $A = 1$: since B is even, we have $B = 0$ and therefore $C = 65$. We find the form $(1, 0, 65)$.

- $A = 2$: then $B = 0$ and $B = -2$ are impossible, so we must have $B = 2$. Now $65 = 2C - 1$ gives $C = 33$, and we get the form $(2, 2, 33)$.
- $A = 3$: clearly $b \neq 0$; $b = \pm 1$ leads to $C = 22$ and to the form $(3, \pm 2, 22)$.
- $A = 4$: this is again impossible since $65 \equiv -b^2 \bmod 4$ is not solvable.
- $A = 5$: For $B = 0$ we find $(5, 0, 13)$. From $65 = 5C - b^2$ we see that b must be divisible by 5, hence B must be divisible by 10, and this only works for $B = 0$.
- $A = 6$: here we find $b = 1$ and $C = 11$, that is, the forms $(6, \pm 2, 11)$. The cases $b = 2$ and $b = 3$ lead to contradictions.
- $A = 7$: this is impossible since $(\frac{-65}{7}) = -1$.
- $A = 8$: this contradicts $65 \equiv -b^2 \bmod 4$.
- $A = 9$: here we check that $65 = 9C - b^2$ for integers b with $|b| \leq 4$ is only solvable for $b = 4$, leading to the form $(9, 8, 9)$.

Thus the set of reduced forms of discriminant $-4 \cdot 65$ is

$$\{(1, 0, 65), (2, 2, 33), (3, \pm 2, 22), (5, 0, 13), (6, \pm 2, 11), (9, 8, 9)\}.$$

Reduced forms have another nice property, which will turn out to be an important tool in various proofs:

Lemma 4.5. *If $Q = (A, B, C)$ is reduced, then the three smallest integers represented by Q are A, C, and $A - |B| + C$.*

Proof. Clearly these integers are represented by Q since $Q(1, 0) = A$, $Q(0, 1) = C$ and $Q(1, \pm 1) = A \pm B + C$.

In order to show that these are the smallest integers represented by Q we have to show that $Q(x, y) \geq A - |B| + C$ for integers x, y with $xy > 1$. We now distinguish three cases:

- $|x| = |y|$. Then $Q(x, y) = x^2(A \pm B + C) \geq (A - |B| + C)x^2 > A - |B| + C$.
- $|x| > |y|$. Then
$$\begin{aligned} Q(x, y) &\geq Ax^2 - |B||xy| + Cy^2 > (A - |B|)|xy| + Cy^2 \\ &\geq (A - |B| + C)y^2 > A - |B| + C. \end{aligned}$$
- $|x| < |y|$. Then $Q(x, y) \geq (A - |B| + C)x^2 > A - |B| + C$.

□

Note that these three integers A, C, and $A - |B| + C$ need not be distinct: if $Q = (1, 1, 1)$, then actually $A = C = A - |B| + C = 1$.

As innocent as this result looks like, it will be the key to our proof that any two equivalent reduced forms are equal, as well as four our discussion of Gauss's conjecture on discriminants with class number 1.

The Reduction Algorithm for Definite Forms

Given a quadratic form (A, B, C) with negative discriminant, how can we find an equivalent reduced form? The algorithm below is a consequence of several simple observations.

First, for $S = \left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)$ we find $Q|_S(x, y) = A(x + sy)^2 + B(x + sy)y + Cy^2$, hence

$$Q|_S = (A, B + 2As, As^2 + Bs + C). \tag{4.2}$$

Thus we can use such a transformation to decrease the size of B while keeping A fixed.

Next, for $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ we get $Q|_S(x, y) = A(-y)^2 + B(-xy) + Cx^2$:

$$Q|_S = (C, -B, A). \tag{4.3}$$

Thus S can be used to exchange A and C.

Here's the algorithm:

input: a primitive quadratic form (A,B,C) with $\Delta < 0$ and $A > 0$.
output: an equivalent reduced form (A'', B'', C'').

1. If $|B| > A$, find $s \in \mathbb{Z}$ with $|B + 2As| \le A$, and put $(A', B', C') = Q|_S$ for $S = \left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)$. Then $|B'| \le A' = A$ (see (4.2)).
2. If $A' \le C'$ goto step 3. If $A' > C'$, use S to replace the form (A', B', C') by $(C', -B', A')$. If $|B'| > C'$, goto step 1.
3. Now we have a quadratic form (A'', B'', C'') with $|B''| \le A'' \le C''$. This form is reduced unless
   (a) $C'' = A''$ and $B'' < 0$; in this case, replace (A'', B'', A'') by the equivalent form $(A'', -B'', A'')$.
   (b) $B'' = -A''$; then replace $(A'', -A'', C'')$ by (A'', A'', C'') using $S = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

This algorithm terminates: in fact, every time the algorithm runs through step 1, the absolute value of the middle coefficient is decreased by at least 1; this clearly can happen only finitely often. Since the output is a reduced form, we have proved

Proposition 4.6. *Every positive definite primitive quadratic form is equivalent to a reduced form.*

Note that this algorithm also can compute the matrix S for which $Q' = Q|_S$: all you have to do is keep track of the matrices used in each step and multiply them together.

Here's an example: start with the form $(A, B, C) = (3, 9, 7)$ with discriminant $\Delta = 8^2 - 4 \cdot 3 \cdot 7 = -3$. From $|9 + 6b| \le 3$ we find that we may take $b = -1$ or $b = -2$. With $b = -2$ we get $(A', B', C') = (3, -3, 1)$. Since $3 > 1$, we switch and get $(1, 3, 3)$. Now we repeat step 1: we find $|3 + 2b| \le 1$ for $b = -1$, and get $(1, 1, 1)$. Thus $(3, 9, 7) \sim (1, 1, 1)$, and this form is reduced.

We now deduce a couple of consequences of Prop. 4.6.

Corollary 4.7. *A (positive definite) quadratic form representing* 1 *is equivalent to the principal form.*

Proof. Let Q be such a quadratic form. Then Q is equivalent to some reduced form Q', which also represents 1. Since 1 is the smallest natural number represented by Q', Lemma 4.5 implies that $Q' = (A, B, C)$ with $A = 1$. Since Q' is reuced, we must have $|B| \leq |A| = 1$, hence $Q' = (1, 0, C)$ or $Q' = (1, 1, C)$. But these are exactly the principal forms. □

We have already seen that every equivalence class $[Q]$ of a quadratic form (positive definite and primitive as usual, with discriminant Δ) contains a reduced form (because Q is equivalent to some reduced form). Now we will prove that every equivalence class contains *exactly one* reduced form. This will have the important consequence that there are exactly as many equivalence classes of forms as there are reduced forms, or in other words, that the number of equivalence classes is just the class number $h(\Delta)$.

Proposition 4.8. *Every positive definite and primitive quadratic form Q is equivalent to a unique reduced form.*

Proof. We have to show that if $Q = (A, B, C)$ and $Q' = (A', B', C')$ are reduced forms with $Q \sim Q'$, then $Q = Q'$.

First we observe that the smallest natural number represented by Q and Q' is A and A', respectively. Since $Q \sim Q'$, they represent the same integers, hence we must have $A = A'$. Note that $C \geq A$ since Q is reduced; we now distinguish some cases.

1. $C > A$. Since $A = Q(\pm 1, 0)$ is represented exactly twice by Q, it is also represented exactly twice by Q', hence $C' = Q'(0, \pm 1) > A' = A$. Now C is the second smallest integer represented by Q, and therefore also by Q'. Since Q and Q' represent the same integers, we must have $C = C'$. Since $\operatorname{disc} Q = \operatorname{disc} Q'$, we see that $|B| = |B'|$. If we had $B' = -B$, then $(A, B, C) = Q \sim Q' = (A, -B, C)$. Assume that $Q' = Q|_S$ for $S = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Then $A = A' = Ar^2 + Brt + Ct^2$, and since $C > A$, the only solutions of this equation are $r = \pm 1$, $t = 0$. Thus $S = \left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)$ or $S = \left(\begin{smallmatrix} -1 & s \\ 0 & -1 \end{smallmatrix}\right)$, hence $-B = B' = 2As + B$, or $As = -B$. Since $|B| \leq A$, we must have $s = 0$ (and then $B = 0 = -B = B'$) or $s = 1$ (and then $B = -A$, which contradicts the assumption that Q is reduced). Thus we have $Q = Q'$ in all cases considered here.

2. $C = A$. Then $A = Q(\pm 1, 0) = Q(0, \pm 1)$, hence A is represented at least four times by Q, hence also by Q'. But this implies $C' = A$ and therefore $C = C'$. As above this implies $B' = \pm B$. But since $(A, B, A) \sim (A, B', A)$ are reduced, B and B' must be positive, and we get $Q = Q'$.

The proof is now complete. □

Representations by Quadratic Forms

If $Q = (A, B, C)$ is a quadratic form with discriminant $\Delta = B^2 - 4AC$ and if Q represents a prime $p \nmid \Delta$, then $(\frac{\Delta}{p}) = +1$. In fact, $p = Ax^2 + Bxy + Cy^2$ implies $4Ap = 4A^2x^2 + 4ABxy + 4ACy^2 = (2Ax + By)^2 - \Delta y^2$. Reduction modulo p gives $\Delta y^2 \equiv (2Ax + By)^2 \bmod p$. Next $p \nmid y$: otherwise we would also have $p \mid x$ and then $p^2 \mid p$. Thus Δ is congruent to a square mod p, and this proves the claim.

The following is a converse to this result:

Lemma 4.9. *If Δ is a nonsquare discriminant and $(\frac{\Delta}{p}) \neq -1$, then there is a quadratic form $Q = (p, B, C)$ with discriminant Δ.*

Proof. Assume first that $p \nmid \Delta$ is odd. Write $\Delta \equiv B^2 \bmod p$. Since p is odd, we may assume that Δ and B have the same parity (otherwise replace B by $p - B$); then $\Delta \equiv B^2 \bmod 4p$, hence there is an integer C such that $\Delta = B^2 - 4pC$. But then $Q = (p, B, C)$ has the desired properties.

Next assume that $p \mid \Delta$ is odd and write $\Delta = pb$. Then $\Delta \equiv 1 \bmod 4$ implies $p \equiv b \bmod 4$, hence $p - b = 4C$. Now $Q = (p, p, C)$ is a form of discriminant $p^2 - 4pC = p(p - 4C) = pb = \Delta$, and obviously Q represents p.

Now consider $p = 2$. If $2 \mid \Delta$, then $\Delta = 4m$. If m is odd, write $m = 1 - 2C$ and take $(2, 2, C)$. If $m = 2C$ is even, take $(2, 0, C)$. □

It is not necessarily true that the form representing p is primitive; for example, 2 is represented by the non-primitive form $2x^2 + 2y^2$ of discriminant $\Delta = -16$, but it is not represented by a primitive form with $\Delta = -16$.

The discriminants Δ with the property that every form of discriminant Δ is primitive are called fundamental. It is easy to see that a discriminant is fundamental if and only if it cannot be written in the form $\Delta = n^2\Delta'$ for some integer $n > 1$ and a discriminant Δ'. In fact, if (A, B, C) is not primitive and $\gcd(A, B, C) = n > 1$, then writing $A = na$, $B = nb$, $C = nc$ shows that $\Delta = B^2 - 4AC = n^2(b^2 - 4ac) = n^2\Delta'$, where Δ' is the discriminant of (a, b, c).

Lemma 4.10. *A discriminant Δ is fundamental if and only if*

$$\Delta = \begin{cases} 4m & \text{for } m \equiv 2, 3 \bmod 4, \\ m & \text{for } m \equiv 1 \bmod 4 \end{cases}$$

with m squarefree.

Proof. Clearly $\Delta = 4m$ is fundamental in the first case since $\Delta' = m$ is not a discriminant (discriminants are $\equiv 0, 1 \bmod 4$); in the second case this is completely obvious.

Assume therefore that Δ is fundamental. If p is a prime with $p^2 \mid \Delta$, then $p = 2$ since otherwise $\Delta = p^2\Delta'$ for some discriminant Δ'. If $\Delta = 4m$, then $4 \nmid m$ since otherwise m is a discriminant; moreover $m \not\equiv 1 \bmod 4$ for the same reason. □

Now observe that $Q = (p, A, C)$ represents p since $p = Q(1, 0)$. Since Q and $Q|_M$ represent the same numbers, p is also represented by $Q|_M$. In particular, p is represented by some reduced form of discriminant Δ.

Corollary 4.11. *If $(\frac{\Delta}{p}) \neq -1$ for some prime p, then p is represented by some reduced form of discriminant Δ.*

Here are some examples that show the power of this result.

If $\Delta = -4$, there is only one reduced form, and we conclude that primes $p \equiv 1 \bmod 4$ have the form $p = x^2 + y^2$.

If $\Delta = -3$, the only reduced form is $x^2 + xy + y^2$, hence every prime $p \equiv 1 \bmod 3$ has the form $p = x^2 + xy + y^2$.

If $\Delta = -20$, there are two reduced forms, namely x^2+5y^2 and $2x^2+2xy+3y^2$. Every prime $p \equiv 1, 3, 7, 9 \bmod 20$ (these are exactly the odd primes $p \nmid 20$ with $(\frac{-5}{p}) = +1$) is represented by one of these forms. We can even say exactly which primes are represented by each of these forms: if $p = x^2 + 5y^2$, then $p \equiv x^2 + y^2 \equiv 1 \bmod 4$, and if $p = 2x^2 + 2xy + 3y^2$, then y is odd, hence $p = 2x(x + y) + 3y^2 \equiv 3 \bmod 4$ because $x(x + y)$ is even and $y^2 \equiv 1 \bmod 4$. Thus the primes $p \equiv 1, 9 \bmod 20$ are represented by the principal form x^2+5y^2, whereas the primes $p \equiv 3, 7 \bmod 20$ are represented by $2x^2 + 2xy + 3y^2$.

Let me recall how we proved these results: if p is a prime with $(\frac{\Delta}{p}) = +1$, then p is represented by some quadratic form of discriminant Δ. Since equivalent forms represent the same integers and have the same discriminant, p is also represented by some reduced form of discriminant Δ.

4.3 Composition

Euler once conjectured that for primes $p \equiv 3, 7 \bmod 20$, we always have $2p = x^2 + 5y^2$ for some $x, y \in \mathbb{Z}$. This is now easy to see: write $p = 2u^2 + 2uv + 3v^2$; then $2p = 4u^2 + 4uv + 6v^2 = (2u + v)^2 + 5v^2$.

Fermat had conjectured before that the product of two distinct primes $p, q \equiv 3, 7 \bmod 20$ is represented by $x^2 + 5y^2$. This is a consequence of the identity

$$(2r^2 + 2rs + 3s^2)(2t^2 + 2tu + 2u^2) = x^2 + 5y^2, \tag{4.4}$$

where

$$x = 2rt + st + ru - 2su, \quad y = ru + su + st.$$

Legendre explained these identities as follows: we have $2(2r^2 + 2rs + 3s^2) = (2r + s)^2 + 5s^2$ for primes $p \equiv 3, 7 \bmod 20$), so multiplying (4.4) through by 4 we get

$$((2r + s)^2 + 5s^2)((2t + u)^2 + 5u^2) = 4(x^2 + 5y^2),$$

which in turn is easily derived from the multiplication

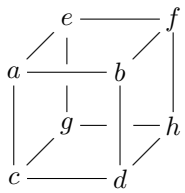
$$\begin{aligned}(2r + s + s\sqrt{-5}\,)(2t + u + u\sqrt{-5}\,)& \\ &= (2r + s)(2t + u) - 5su + [(2r + s)u + (2t + u)s]\sqrt{-5} \\ &= 2(2rt + st + ru - 2su) + 2(ru + su + st)\sqrt{-5}\end{aligned}$$

of complex integers.

The identity (4.4) is a special case of composition, which in turn is some kind of multiplication of equivalence classes of forms of the same discriminant. Composition of forms has always been looked upon as some horribly technical way of defining a group structure on the set of equivalence classes of forms.

While it is true that Gauss's original account of composition was extremely technical and difficult, subsequent work by Dirichlet, Cayley, Dedekind, Weber, Speiser, Riss, Shanks, and – in the last few years – Bhargava have led to a quite simple description of composition.

For integers a, b, c, d, e, f, g, h with $\gcd(a, b, \ldots, h) = 1$ we now consider the cube



.

Each such cube can be sliced in three different ways, producing three pairs of 2×2-matrices (front-back, left-right, up-down):

$$\begin{array}{lll} FB & M_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, & N_1 = \begin{pmatrix} e & f \\ g & h \end{pmatrix}, \\ LR & M_2 = \begin{pmatrix} a & c \\ e & g \end{pmatrix}, & N_2 = \begin{pmatrix} b & d \\ f & h \end{pmatrix}, \\ UD & M_3 = \begin{pmatrix} a & e \\ b & f \end{pmatrix}, & N_3 = \begin{pmatrix} c & g \\ d & h \end{pmatrix}. \end{array}$$

To each cube A we can associate three binary quadratic forms $Q_i = Q_i^A$ by putting

$$Q_i(x, y) = -\det(M_i x + N_i y).$$

This way we find

$$\begin{aligned} Q_1(x, y) &= (bc - ad)x^2 + (bg + cf - ah - de)xy + (fg - eh)y^2, \\ Q_2(x, y) &= (ce - ag)x^2 + (cf + de - ah - bg)xy + (df - bh)y^2, \\ Q_3(x, y) &= (be - af)x^2 + (bg + de - ah - cf)xy + (dg - ch)y^2. \end{aligned}$$

Setting $Q_i = (A_i, B_i, C_i)$ we find that in the FB-slicing we have $A_1 = -\det F$ and $C_1 = -\det B$, where F and B denote the matrices forming the front and the back face of the cube. Similarly we have $A_2 = -\det L$ and $C_2 = -\det R$ in the LR-slicing. The matrices F and L have the edge ac in common; the diagonal matrix D_{FB} satisfies $\frac{1}{2}(B_1 + B_2) = -\det D_{FB}$.

A simple calculation shows that $\operatorname{disc} Q_1 = \operatorname{disc} Q_2 = \operatorname{disc} Q_3$ is equal to

$$\begin{aligned} \operatorname{disc}(A) := {} & a^2h^2 + b^2g^2 + c^2f^2 + d^2e^2 \\ & - 2(abgh + cdef + acfh + bdeg + aedh + bfcg) + 4(adfg + bceh). \end{aligned}$$

The Action of $\mathrm{SL}_2(\mathbb{Z})$ on Cubes

We now define an action of $S = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ on the cube by replacing the cube with front M_1 and back N_1 by the cube with front $rM_1 + tN_1$ and back $sM_1 + uN_1$.

Lemma 4.12. *Let A be a cube, $S \in \mathrm{SL}_2(\mathbb{Z})$, and let $A' = A|_S$ be the cube we get by letting S act on A; then $\operatorname{disc} A' = \operatorname{disc} A$. If the associated quadratic forms are denoted by Q_i and Q_i', then $Q_1' = Q_1|_S$, $Q_2' = Q_2$, and $Q_3' = Q_3$.*

Proof. We know that $Q_1 = -\det(M_1 x + N_1 y)$; applying $S = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right)$ we see that

$$\begin{aligned} Q_1'(x,y) &= -\det\big((rM_1 + tN_1)x + (sM_1 + uN_1)y\big) \\ &= -\det\big(M_1(rx+sy) + N_1(tx+uy)\big). \end{aligned}$$

Since $Q_1 = (A,B,C) = -\det(M_1 x + N_1 y)$, we find

$$\begin{aligned} Q_1'(x,y) &= A(rx+sy)^2 + B(rx+sy)(tx+uy) + C(tx+uy)^2 \\ &= (A',B',C') \end{aligned}$$

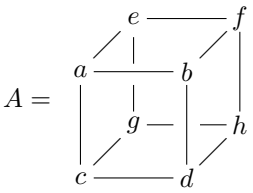
for

$$\begin{aligned} A' &= Ar^2 + Brt + Ct^2, \\ B' &= 2(Ars + Ctu) + B(ru+st), \\ C' &= As^2 + Bsu + Cu^2. \end{aligned}$$

Thus we see that $Q_1'(x,y) = Q|_S(x,y)$ as claimed.

Observe also that the action of $\mathrm{SL}_2(\mathbb{Z})$ is trivial on the quadratic forms Q_2 and Q_3, since this group acts by row and column operations on M_j and N_j for $j = 2, 3$, hence does not change the determinant $\det(M_j x + N_j y)$. □

The last claim can be made more obvious by representing the cube

$A =$ 

by the 2×4-matrix

$$M(A) = \begin{pmatrix} a & b & c & d \\ e & f & g & h \end{pmatrix}.$$

Then the cube $A|_S$ is represented by the matrix

$$M(A|_S) = \begin{pmatrix} ra+te & rb+tf & rc+tg & rd+th \\ sa+ue & sb+uf & sc+ug & sd+uh \end{pmatrix} = S^t M(A).$$

Note that

$$M(A_{ST}) = (ST)^t M(A) = T^t S^t M(A) = T^t M(A|_S) = M((A|_S)|_T),$$

so this is indeed an action.

It is now easily checked that the six minors of $M(A)$ essentially are the coefficients of Q_2^A and Q_3^A; for example, we have $ag - ce = \det M_2 = -A_2$ etc. Since the minors of $M(A|_S)$ are the minors of $M(A)$ multiplied by S^t from the left, their determinants are the same, showing that the forms Q_2 and Q_3 computed from A and $A|_S$ are indeed the same.

Now instead of letting $\mathrm{SL}_2(\mathbb{Z})$ act on the pair (M_1, N_1) as above we can also let it act on (M_2, N_2) and (M_3, N_3). In this way we get an action of the group $\Gamma = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ on the set of cubes; the action $A|_S$ described above now is $A|_{(S,I,I)}$, where I is the identity element in $\mathrm{SL}_2(\mathbb{Z})$. Note that the action of the three factors in Γ commutes: if you let an element (S_1, S_2, S_3) act on a cube then it does not matter whether you first let S_1 act on (M_1, N_1) and then S_2 on (M_2, N_2) or the other way round (check this!).

Observe also that the action of the subgroup $I \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ of Γ is trivial on the quadratic form Q_1, since this subgroup acts by row and column operations on M_1 and N_1, hence does not change the determinant $\det(M_i x + N_i y)$.

The Group Law

Now let Δ be a nonsquare discriminant (there is no need to exclude positive discriminants here), and write $\Delta = \sigma^2 - 4m$ for $\sigma \in \{0, 1\}$. We will make the set of equivalence classes of primitive binary quadratic forms of discriminant Δ into a group whose neutral element is the class $1 = [I]$ of the principal form $I = (1, \sigma, m)$.

To this end, let Q_1 and Q_2 be two such forms. We will prove below that there always exists a cube A with $Q_1^A = Q_1$ and $Q_2^A = Q_2$. Put $Q_3 = Q_3^A$. Then we say that $[Q_1][Q_2][Q_3] = 1$.

We now sketch a proof that this does indeed define a group law. Before we can check the axioms we have to show that any two form classes $[Q_1]$ and $[Q_2]$ can be composed, i.e., that there is a cube A with $Q_1^A = Q_1$ and $Q_2^A = Q_2$. We will postpone this and first check the group axioms.

Composition is Well Defined

There are various things to show here:

1. If $\mathbb{Q}_1 \sim Q_1'$ and $Q_2 \sim Q_2'$, then $[Q_1][Q_2] = [Q_1'][Q_2']$.
2. If A and B are cubes with $Q_1^A = Q_1^B$ and $Q_2^A = Q_2^B$, then we have $Q_3^A \sim Q_3^B$.

To prove the first claim, assume that $Q_1' = Q_1|_S$ and $Q_2' = Q_2|_T$, let A be a cube with $Q_1^A = Q_1$ and $Q_2^A = Q_2$, and put $Q_3 = Q_3^A$. Then $B = A|_{(S,T,I)}$ is

a cube with $Q_1^B = Q_1'$, $Q_2^B = Q_2'$, and $Q_3^B = Q_3$. The definition of composition applied to the cube A shows $[Q_1][Q_2][Q_3] = [I]$, and similarly B shows that $[Q_1'][Q_2'][Q_3] = [I]$. In particular, $[Q_1][Q_2] = [Q_3]^{-1} = [Q_1'][Q_2']$ (the existence of inverses will also be proved below).

The second claim is proved by invoking Gauss's Lemma (there are actually three results known as Gauss's Lemma: one in the theory of quadratic residues, one in the theory of polynomial rings, and the following):

Lemma 4.13 (Gauss's Lemma)**.** *Let*

$$M = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ q_1 & q_2 & \cdots & q_n \end{pmatrix} \quad \textit{and} \quad M' = \begin{pmatrix} p_1' & p_2' & \cdots & p_n' \\ q_1' & q_2' & \cdots & q_n' \end{pmatrix}$$

be two matrices with the following properties:

1. *the minors of M are coprime;*
2. *there is an integer k such that each minor of M' is k times the corresponding minor of M.*

Then there is a matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with determinant k such that $M' = AM$.

Proof. Since the minors of M are coprime, there exist n^2 integers x_{ik} such that

$$\sum_{i,k=1}^{n} x_{ik} \begin{vmatrix} p_i & p_k \\ q_i & q_k \end{vmatrix} = 1.$$

Then we find

$$\begin{aligned} p_k \begin{vmatrix} p_i' & p_j' \\ q_i & q_j \end{vmatrix} + q_k \begin{vmatrix} p_i & p_j \\ p_i' & p_j' \end{vmatrix} &= p_i' \begin{vmatrix} p_k & p_j \\ q_k & q_j \end{vmatrix} - p_j' \begin{vmatrix} p_k & p_i \\ q_k & q_i \end{vmatrix} \\ &= \frac{1}{k} \left(p_i' \begin{vmatrix} p_k' & p_j' \\ q_k' & q_j' \end{vmatrix} - p_j' \begin{vmatrix} p_k' & p_i' \\ q_k' & q_i' \end{vmatrix} \right) = \frac{1}{k} p_k' \begin{vmatrix} p_i' & p_j' \\ q_i' & q_j' \end{vmatrix} = p_k' \begin{vmatrix} p_i & p_j \\ q_i & q_j \end{vmatrix}. \end{aligned}$$

Now set

$$\begin{aligned} a &= \sum x_{ij} \begin{vmatrix} p_i' & p_j' \\ q_i & q_j \end{vmatrix}, & b &= \sum x_{ij} \begin{vmatrix} p_i & p_j \\ p_i' & p_j' \end{vmatrix}, \\ c &= \sum x_{ij} \begin{vmatrix} q_i' & q_j' \\ q_i & q_j \end{vmatrix}, & d &= \sum x_{ij} \begin{vmatrix} p_i & p_j \\ q_i' & q_j' \end{vmatrix}. \end{aligned}$$

Then a, b, c, d are integers satisfying

$$ap_k + bq_k = \sum x_{ij} \left(p_k \begin{vmatrix} p_i' & p_j' \\ q_i & q_j \end{vmatrix} + q_k \begin{vmatrix} p_i & p_j \\ p_i' & p_j' \end{vmatrix} \right) = p_k' \sum x_{ij} \begin{vmatrix} p_i & p_j \\ q_i & q_j \end{vmatrix} \quad = \quad p_k'.$$

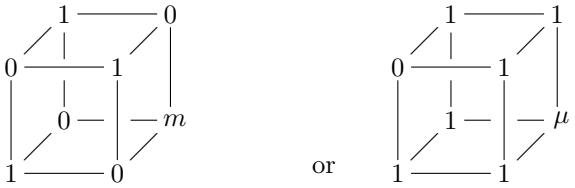
Similarly, we find $cp_k + dq_k = q_k'$, and this completes the proof. □

In the applications we have in mind, the matrix will always be a 2×4-matrix, and we always will have $k = 1$.

Assume now that A and B are cubes with $Q_2^A = Q_2^B = Q_2$ and $Q_3^A = Q_3^B = Q_3$ (we have changed indices); we have to show that $Q_1^A \sim Q_1^B$. Let $M(A)$ and $M(B)$ denote the 2×4-matrices corresponding to A and B; then the six minors of $M(A)$ and $M(B)$ are determined by the coefficients of Q_2^A, Q_3^A and Q_2^B, Q_3^B, respectively, so they must be equal. Gauss's Lemma then says that there is some $S \in \mathrm{SL}_2(\mathbb{Z})$ such that $M(B) = S^t M(A)$. But then $Q_1^B = Q_1^A|_S \sim Q_1^A$ as claimed.

Neutral Element

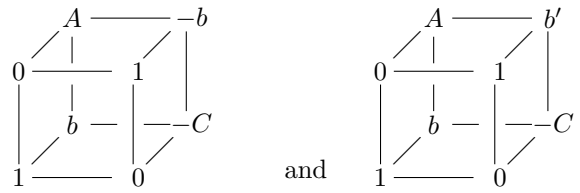
The first axiom to check is that $[I][I][I] = [I]$ for the principal form $I = (1, \sigma, m)$. To this end we have to come up with a cube whose three quadratic forms are all equivalent to I. This is easy: just take



according as $\Delta = 4m$ or $\Delta = 4m + 1 = 4\mu - 3$, with $\mu = m + 1$. Note that these cubes are "triply symmetric": rotation by 120^o about the long diagonal containing m and μ, respectively, leaves the cubes invariant.

Inverse Elements

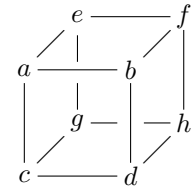
The inverse element of $[Q]$, where $Q = (A, B, C)$, is the class of the form $Q^- = [A, -B, C]$. We know that $B \equiv \Delta \bmod 2$; thus we can put $B = 2b$ if $\Delta = 4m$, and $B = 2b - 1$ if $\Delta = 1 + 4m$. With $b' = 1 - b$ we then find that the two cubes

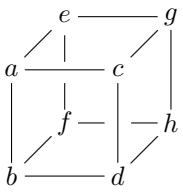


give rise to the quadratic forms $Q_1 = I$, $Q_2 = (A, B, C)$, and $Q_3 = (A, -B, C)$. This implies that $[I][Q][Q^-] = [I]$, and now the claim follows.

Commutativity

Next we show that composition is abelian, i.e., that $[Q_1][Q_2] = [Q_2][Q_1]$. To see this, we only have to observe that the quadratic forms attached to the cubes

$A =$ 

$B =$ 

satisfy $Q_1^B = Q_1^A$, $Q_2^B = Q_3^A$, and $Q_3^B = Q_2^A$.

Associativity

This is the axiom that is the most difficult to check: given classes $[Q_1]$, $[Q_2]$, $[Q_3]$ of discriminant Δ, we have to show that $([Q_1][Q_2])[Q_3] = [Q_1]([Q_2][Q_3])$.

The "textbook version" proceeds by replacing the Q_i by certain equivalent forms Q_i' in such a way that composition becomes more or less obvious (see "Dirichlet composition" below). So far I haven't found a simple direct proof of associativity using Bhargava's cubes. If you can come up with one, let me know.

Dirichlet composition

The basic idea is the following: since we are only interested in the equivalence class of the composition of $[Q_1]$ and $[Q_2]$, we may use the action of $\mathrm{SL}_2(\mathbb{Z})$ to replace Q_1 and Q_2 by equivalent forms before we compose them.

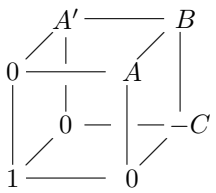
Let us now call quadratic forms (A, B, C) and (A', B', C') with nonsquare discriminant Δ *concordant* if the coefficients have the following properties:

1. $B = B'$;
2. $A' \mid C$ and $A \mid C'$.

The composition of concordant forms turns out to be extremely simple:

Proposition 4.14. *If Q and Q' are concordant, then $Q = (A, B, A'C)$ and $Q' = (A', B, AC)$ for integers A, A', B, C, and we have $[Q] \oplus [Q'] = [Q'']$ for $Q'' = (AA', B, C)$.*

Proof. Consider the cube A given by



Its associated quadratic forms are

$$\begin{aligned} Q_1 &= Ax^2 + Bxy + A'Cy^2, \\ Q_2 &= A'x^2 + Bxy + ACy^2, \\ Q_3 &= AA'x^2 - Bxy + Cy^2. \end{aligned}$$

This implies

$$[Q_1][Q_2] = [Q_3]^{-1} = [(AA', B, C)],$$

which is what we had to prove. □

Example 1. Consider the forms $Q_1 = (2, 2, 33)$ and $Q_2 = (3, 2, 22)$. Since these are concordant, we find $[Q_1][Q_2] = [Q_3]$ with $Q_3 = (6, 2, 11)$.

Example 2. Consider the form $Q = (2, 1, 2)$ with discriminant $\Delta = -15$. Here $[Q]^2 = [Q][Q] = [Q']$ for $Q' = (4, 1, C'')$, and $-15 = 1^2 - 4 \cdot 4C''$ gives $C'' = 1$. Thus $[Q]^2 = [(4, 1, 1)]$, and since $(4, 1, 1) \sim (1, 1, 4)$, $[Q]$ has order 2 in $\mathrm{Cl}\,(-15)$.

Example 3. The forms (A, B, C) and (C, B, A) are concordant, and we find $[(A, B, C)][(C, B, A)] = [(AC, B, 1)] = [(1, -B, AC)]$.

Example 4. Consider the form $Q = (2, 1, 3)$ with discriminant $\Delta = -23$. Here Q and Q are not concordant, so we have to replace them by equivalent forms before we can compose them. Now $Q' = Q|_M = (2, -3, 4)$ for $M = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, hence $[Q]^2 = [Q']^2 = [Q'']$ for $Q'' = (4, -3, 2)$. Now $(4, -3, 2) \sim (2, 3, 4) \sim (2, -1, 3)$.

By Example 3, we also have $3[Q] = [(2, 1, 3)][(2, -1, 3)] = [(2, 1, 3)][(3, 1, 2)] = [(6, 1, 1)] = [1, -1, 6] = [1, 1, 6]$.

In order to be able to actually work with Dirichlet composition, we need a method for changing two given forms Q_1, Q_2 into equivalent forms $Q_1' \sim Q_1$ and $Q_2' \sim Q_2$ such that Q_1' and Q_2' are concordant. We start with a simple

Lemma 4.15. *Let $Q = (A, B, C)$ be a primitive quadratic form. Then for any $N \in \mathbb{N}$ there are $r, s \in \mathbb{Z}$ such that $Q(r, s)$ is coprime to N.*

Proof. Write $N = rst$, where $(r, C) = 1$, and where the primes $p \mid s$ and $q \mid t$ satisfy $p \mid C$, $p \nmid A$, $q \mid A$ and $q \mid C$. Then we find

$$\begin{aligned} \gcd(Q(r, s), r) &= \gcd(Cs^2, r) = 1, \\ \gcd(Q(r, s), s) &= \gcd(Ar^2, s) = 1, \\ \gcd(Q(r, s), t) &= \gcd(Brs, t) = 1, \end{aligned}$$

where we have used that $\gcd(B, t) = 1$ because t is primitive. □

Now we can construct concordant forms:

Lemma 4.16. *Let Q_1 and Q_2 be quadratic forms of discriminant Δ. Then for any $N \in \mathbb{N}$ there exist concordant forms $Q_1' = (A, B, C)$ and $Q_2' = (A', B, C')$ equivalent to Q_1 and Q_2, respectively, and such that* $\gcd(A, A') = \gcd(AA', N) = 1$.

Proof. First we show that we can choose a form $R_1 = (A_1, B_1, C_1) \sim Q_1$ with $\gcd(A_1, N) = 1$. In fact, pick $r, s \in \mathbb{Z}$ coprime with $A_1 = Q_1(r, s)$ and $\gcd(A_1, N) = 1$. By Bezout there are $t, u \in \mathbb{Z}$ with $ru - st = 1$; then $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Now $R_1 = Q_1|_M = (A_1, B_1, C_1) \sim Q_1$.

Similarly we choose $R_2 = (A_2, B_2, C_2) \sim Q_2$ with $\gcd(A_2, A_1 N) = 1$.

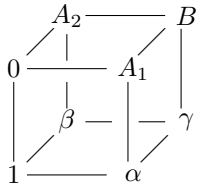
Now we find integers n_1, n_2 such that $B_1 + 2A_1 n_1 = B_2 + 2A_2 n_2$. This equation can be written in the form $A_1 n_1 - A_2 n_2 = (b_1 - b_2)/2$, ans this has an integral solution because $b_1 \equiv \Delta \equiv b_2 \bmod 2$ and $\gcd(A_1, A_2) = 1$. Now put $M_i = \left(\begin{smallmatrix} 1 & n_i \\ 0 & 1 \end{smallmatrix}\right)$ and $B = B_i + 2A_i n_i$; then the forms $Q_i' = R|_{M_i} = (A_j, B, C_j)$ have the desired properties. □

Shanks' Algorithm

We have seen how to attach three binary quadratic forms of the same discriminant to a cube A. Now we show that, conversely, to each pair of primitive binary quadratic forms of the same discriminant Δ we can find a cube A giving rise to these forms.

Proposition 4.17. *Given two primitive forms Q_1, Q_2 of discriminant Δ, there exists a cube A such that $Q_1^A = Q_1$ and $Q_2^A = Q_2$.*

The proof we will give goes back to Shanks. The basic idea is the following: assume that $Q_i = (A_i, B_i, C_i)$, and that $(A_1, A_2) = 1$. Then we form the cube



with $B = \frac{1}{2}(B_1 + B_2)$. Then the determinants of the front face, the left face, and the diagonal have the right values, namely $-A_1$, $-A_2$, and $-B$. The values of α, β and γ can then be determined from the three equations

$$\begin{aligned} B\alpha - A_1\gamma &= C_2, \\ A_2\alpha - A_1\beta &= (B_2 - B_1)/2, \\ B\beta - A_2\gamma &= C_1. \end{aligned}$$

It then remains to show that these equations have a common solution (α, β, γ), and that it is integral.

The first claim is easy to show: multiplying the first and the third equation by A_2 and A_1, respectively, and subtracting them from each other gives

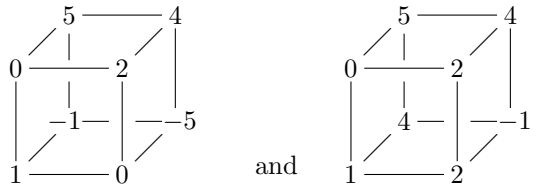
$$B(A_1\beta - A_2\alpha) = A_2 C_2 - A_1 C_2.$$

Now we use the fact that the right hand side equals $B(B_2 - B_1)/2$; dividing through by B then gives the second equation. The integrality of the solutions will be proved below.

Example. Consider the two forms $Q_1 = (2, 2, 21)$ and $Q_2 = (5, 6, 10)$ of discriminant $-4 \cdot 41$. The system of equations becomes

$$\begin{aligned} 4\alpha - 2\gamma &= 10, \\ 5\alpha - 2\beta &= 2, \\ 4\beta - 5\gamma &= 21. \end{aligned}$$

The solutions $(\alpha, \beta, \gamma) = (0, -1, -5)$ and $(2, 4, -1)$ give the cubes



Computing the associated forms gives Q_1 and Q_2 (of course), as well as

$$Q_3 = (10, -6, 5) \quad \text{and} \quad Q_3' = (10, 14, 9).$$

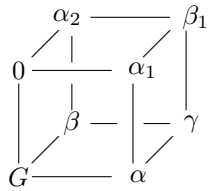
In particular we have $[Q_1][Q_2][Q_3] = 1$, or $[Q_1][Q_2] = [(10, 6, 5)] = [(5, -6, 10)]$. This shows that $BC = C^-$.

Note that you can get the second cube from the first by adding the top face to the bottom. This proves in particular that $Q_3 \sim Q_3'$.

Of course we know that there must be infinitely many solutions since we can make $\mathrm{SL}_2(\mathbb{Z})$ act on the cube in such a way that the two forms Q_1 and Q_2 are not changed.

In general, the condition $(A_1, A_2) = 1$ will not be satisfied, and then we will have to work a little bit harder. We could, of course, force $(A_1, A_2) = 1$ by changing e.g. Q_2 using the action of $\mathrm{SL}_2(\mathbb{Z})$; here, we will follows Shanks' lead and attack composition head-on.

Proof. Let $G = \gcd(A_1, A_2, B)$, put $\alpha_i = A_i/G$ and $\beta_1 = B/G$, and form the cube



Now write $\gcd(A_1, B) = GH$; from

$$\frac{B_2 + B_1}{2} \frac{B_2 - B_1}{2} = A_2 C_2 - A_1 C_1$$

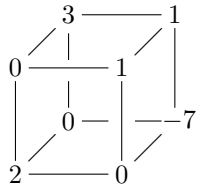
we deduce that $GH \mid A_2C_2$. Since $\gcd(GH, A_2) = G$, we must have $H \mid C_2$. With $A_1 = GHa_1$, $B = GHb$, and $C_2 = Hc_2$, $\beta_1\alpha - \alpha_1\gamma = C_2$ becomes $b\alpha - a_1\gamma = c_2$, where $\gcd(a_1, b) = 1$. Thus the equation $b\alpha - a_1\gamma = c_2$ has an integral solution (α, γ).

Now look at the last equation $\beta_1\beta\alpha_2\gamma = C_1$, i.e., $Hb\beta - \alpha_2 G\gamma = C_1$. We claim that β is an integer, i.e., that $\frac{A_2}{G}\gamma \equiv C_1 \bmod b$. We know that $a_1\gamma \equiv -c_2 \bmod b$.

We know $A_2C_2 \equiv A_1C_1 \bmod B$, hence $\frac{A_2}{G}Hc_2 \equiv a_1HC_1 \bmod Hb$, and therefore $\frac{A_2}{G}c_2 \equiv a_1C_1 \bmod b$.

$a_1\frac{A_2}{G}\gamma - a_1C_1 \equiv a_1\frac{A_2}{G}\gamma - \frac{A_2}{G}c_2 \equiv \frac{A_2}{G}(a_1\gamma - c_2) \bmod Hb$. □

Example. Consider the forms $Q_1 = (2, 2, 21)$ and $Q_2 = (6, 2, 7)$ of discriminant $-4\cdot 41$. Here $G = \gcd(A_1, A_2, B) = 2$, $\gcd(A_1, B) = \gcd(2, 2) = 2$, hence $H = 1$. We get the system of equations $\alpha - \gamma = 7$ and $\beta - 3\gamma = 21$, which has the solution $\gamma = -7$ and $\alpha = \beta = 0$. Thus we get the cube



with the associated forms Q_1, Q_2 and $Q_3 = (3, -2, 14)$. Thus $[Q_1][Q_2] = [(3, 2, 14)]$.

Exercises

4.1 Determine all positive definite quadratic forms Q with the property that the smallest integers represented by Q are 2, 3, and 5.

4.2 Determine the class number $h(-23)$.

4.3 Compute $h(-31)$.

4.4 Compute $h(-43)$.

4.5 Reduce the form $(101, 20, 1)$.

4.6 Reduce the form $(3, 4, 3)$.

4.7 Reduce the forms $(5, 16, 21)$ and $(7, 16, 15)$.

Chapter 5

Gauss's Class Number 1 Problem

It was conjectured by Gauss and proved by Heegner, Baker, and Stark (independently in 1952, 1966, and 1967) that the only negative fundamental discriminants (discriminants Δ with $\Delta \equiv 1 \bmod 4$ squarefree, or $\Delta/4 \equiv 3 \bmod 4$ squarefree) with class number 1 are

$$\Delta = -3, -4, -7, -8, -11, -19, -43, -67, -163.$$

Actually Gauss's original conjecture looked somewhat different since he only worked with quadratic forms with even middle coefficient.

The proof of this conjecture is well beyond our means; on the other hand, there are some special cases that are accessible, and at the very least we can give some reasons that explain why everybody (including Gauss) believed the conjecture was true.

5.1 Gauss's Conjecture

We can classify the fundamental discriminants into three classes: those of the form $\Delta = 4m$ for squarefree $m \equiv 2, 3 \bmod 4$ and the squarefree $\Delta = 1 - 4m$. The first case is actually easily taken care of:

Proposition 5.1. *Assume that $m \equiv 2, 3 \bmod 4$ is squarefree and $m < -2$, and let $\Delta = 4m$. Then $h(\Delta) > 1$.*

Thus the only two discriminants of this form with class number 1 are $\Delta = -4$ and $\Delta = -8$.

Proof. Consider the case $m \equiv 3 \bmod 4$; then $(1, 0, -m)$ and $Q = (2, 2, \frac{1-m}{2})$ are reduced forms of discriminant $4m$, hence $h(\Delta) \geq 2$.

Next let $m \equiv 2 \bmod 4$. If $m < -2$, then $m = 2a$ for some odd integer a, and $(1, 0, -m)$ and $(2, 0, -a)$ are reduced forms of discriminant $\Delta = -4m$. Thus again $h(\Delta) \geq 2$. □

In oder to solve Gauss's conjecture, it is therefore "only" necessary to consider odd discriminants Δ.

Lemma 5.2. *If $\Delta = 1 - 4m$ and $h(\Delta) = 1$, then Δ is prime.*

Proof. If $p \mid \Delta$, then p is represented by some reduced form of discriminant Δ. Since $h(\Delta) = 1$, it is represented by the principal form. Since $\Delta = 1 - 4m$, this means $p = x^2 + xy + my^2$. But then $4p = (2x+1)^2 - \Delta y^2$ shows that $p \mid (2x+1)$. If $|2x + 1| \geq p$, then $(2x + 1)^2 - \Delta y^2 \geq p^2 + |\Delta| > p$, hence $2x + 1 = 0$. But then $4p = \Delta y^2$ and the fact that Δ is odd imply $y = \pm 2$, hence $p = \Delta$. □

Lemma 5.3. *If $\Delta = 1 - 4m$ and $h(\Delta) = 1$, then $(\frac{\Delta}{p}) = -1$ for all $p < m$.*

Since $(\frac{\Delta}{m}) = (\frac{1-4m}{m}) = +1$, this is best possible. If we define $(\frac{\Delta}{2}) := (\frac{2}{\Delta})$ for discriminants $\Delta \equiv 1 \bmod 4$, the result above also holds for $p = 2$.

Proof. If $\Delta \equiv 1 \bmod 8$, write $\Delta = 1 - 8m$; then $Q = (2, 1, m)$ has discriminant Δ, and Q represents 2. Since $h(\Delta) = 1$, we see that 2 is also represented by the form $(1, 1, m)$. This is only possible if $m = 2$, i.e., if $\Delta = -7$ (in this case the condition $(\frac{\Delta}{p}) = -1$ for all $p < m$ is vacuously true). Thus $\Delta \equiv 5 \bmod 8$, i.e., $(\frac{\Delta}{2}) = -1$, for all $\Delta < -7$.

If $p \mid \Delta$ or $(\frac{\Delta}{p}) = +1$, then p is represented by $(1, 1, m)$, hence $m \leq p$. □

Note that, for $\Delta = -163$, we have $(\frac{\Delta}{p}) = -1$ for $p = 2, 3, 5, 7, \ldots, 37$! This already explains why there are so few discriminants with class number 1.

The converse of Lemma 5.3 is also true, even in a stronger form:

Lemma 5.4. *If $\Delta = 1-4m$ and $(\frac{\Delta}{p}) = -1$ for all $p < \sqrt{-\Delta/3}$, then $h(\Delta) = 1$.*

Proof. Let $Q = (A, B, C)$ be a reduced form with discriminant Δ. If $A > 1$, then there is a prime $p \mid A$. From $\Delta = B^2 - 4AC \equiv B^2 \bmod p$ we deduce that $p \mid \Delta$ or $(\frac{\Delta}{p}) = +1$. Since there are no such primes $< \sqrt{-\Delta/3}$, we must have $A > \sqrt{-\Delta/3}$; but this contradicts the assumption that Q be reduced.

Thus the only reduced form must have $A = 1$, hence is equal to $(1, 1, m)$, and we conclude that $h(\Delta) = 1$. □

This proves the following

Theorem 5.5. *Let $\Delta = 1 - 4m$ be squarefree and negative. Then the following statements are equivalent:*

1. *$h(\Delta) = 1$;*
2. *$(\frac{\Delta}{p}) = -1$ for all $p < \sqrt{-\Delta/3}$;*
3. *$(\frac{\Delta}{p}) = -1$ for all $p < m$.*

5.2 Euler's Prime Producing Polynomials

Consider the numbers

$$41,\ 41+2,\ 41+2+4,\ 41+2+4+6,\ \ldots.$$

These are the values of the polynomial $f(x) = x^2+x+41$ for $x \in \mathbb{N}$. Computing them we get the numbers $41, 43, 47, 53, 61, 71, 83, 97, 113, 131, 151, 173, 197, 223$, 251, 281, 313, 347, 383, 421, 461, 503, 547, 593, 641, 691, 743, 797, 853, 911, 971, 1033, $1097, 1163, 1231, 1301, 1373, 1447, 1523, 1601$. All of these numbers are prime! In fact, Euler discovered in 1772 that the polynomial $f(x) = x^2+x+41$ attains only prime values for $x = 0, 1, 2, \ldots, 39$ (note that $f(40) = 40^2+40+41 = 40 \cdot 41 + 41 = 41^2$ is composite).

More generally, the polynomials $f(x) = x^2 + x + m$ for $m = 3, 5, 11, 17, 41$ yield prime values for all $x = 0, 1, \ldots, m-2$. The discriminant of $x^2 - x + m$ is $1 - 4m$, which equals -11, -19, -43, -67 and -163 for the above values of m. This is of course no accident.

In fact, $f(x) = Q(x, 1)$ for the principal quadratic form $Q = (1, 1, m)$ of discriminant $\Delta = 1 - 4m$.

Lemma 5.6. *If $\Delta = 1-4m$ and $h(\Delta) = 1$, and if the prime q divides a number properly represented by $Q = (1, 1, m)$ (this means that $q \mid Q(x, y)$ for coprime integers $x, y \in \mathbb{Z}$), then q is represented by Q.*

Proof. From $x^2 + xy + my^2 \equiv 0 \bmod q$ we find $(2x + 1)^2 + \Delta y^2 \equiv 0 \bmod q$. If $q \mid y$, then $y \mid (2x + 1)$ and $q^2 \mid q$: contradiction. Thus $\Delta \equiv (\frac{2x+1}{y})^2 \bmod q$, or $(\frac{\Delta}{q}) = +1$. Thus q is represented by some reduced form of discriminant Δ, and since $h(\Delta) = 1$, it is represented by Q. □

Theorem 5.7. *If $h(\Delta) = 1$ for some fundamental discriminant $\Delta = 1-4m$ with $m > 2$, then $f(x) = x^2+x+m$ attains only prime values for $x = 0, 1, \ldots, m-2$.*

The converse is actually also true.

Proof. Observe that $f(x) = Q(x, 1)$ for $Q = (1, 1, m)$. Assume that $f(r) = p_1 \cdots p_t$ is composite ($p_i > 1$); then the p_i are also represented by Q. Since the smallest integer > 1 represented by Q is m, we have $p_i \geq m$. Since $t > 1$, we must have $f(r) \geq m^2$, i.e., $r^2 + r \geq m^2 - m$. This implies $(2r+1)^2 \geq (2m-1)^2$ and therefore $r \geq m - 1$. □

Here is another curious fact:

d	$\exp(\pi\sqrt{-d})$
-43	884736743.99977746603490666...
-67	147197952743.99999866245422450...
-163	262537412640768743.99999999999925007...

As you can see, the values of $e^{\pi\sqrt{-d}}$ is very close to an integer for these values of d. This phenomenon becomes even more visible if we subtract 744 and take the cube root:

d	$(\exp(\pi\sqrt{-d}) - 744)^{1/3}$
-11	$31.99809333222744098975227354\ldots$
-19	$95.99999195891694508468060476\ldots$
-43	$959.99999999991951173137537734\ldots$
-67	$5279.99999999999998400738235224\ldots$
-163	$640319.99999999999999999999999939\ldots$

What is even more amazing is the fact that the integer x approximated by $\sqrt[3]{e^{\pi\sqrt{-d}} - 744}$, satisfies the diophantine equation $x^3 + 1728 = -dy^2$ for some $y \in \mathbb{N}$. Look and see:

$$\begin{aligned} 11 \cdot 56^2 &= 32^3 + 1728 \\ 19 \cdot 216^2 &= 96^3 + 1728 \\ 43 \cdot 4536^2 &= 960^3 + 1728 \\ 67 \cdot 4\,6872^2 &= 5280^3 + 1728 \\ 163 \cdot 401\,33016^2 &= 640320^3 + 1728 \end{aligned}$$

Unfortunately, explaining these facts requires more advanced techniques (modular forms, elliptic curves, complex multiplication, ...); these are in fact exactly the techniques one needs to give Heegner's (or Stark's) proof of Gauss's conjecture.

5.3 Quadratic Number Fields

Let m be a squarefree integer, and put

$$\omega = \begin{cases} \sqrt{m} & \text{if } m \equiv 2, 3, \bmod 4, \\ \frac{1+\sqrt{m}}{2} & \text{if } m \equiv 1 \bmod 4. \end{cases}$$

Consider the set

$$\mathcal{O}_\Delta = \mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}\}.$$

This is a ring: its elements can be added, subtracted, and multiplied in the obvious way. Note that, in the case $m = 1+4n$, we have $\omega^2 = \frac{1}{4}(1+m+2\sqrt{m}\,) = n + \omega$. Examples of such rings are the ring $\mathbb{Z}[i]$ of Gaussian integers, whose elements are of the form $a + bi$, or the ring $\mathbb{Z}[\rho]$ of Eisenstein integers, where $\rho = \frac{-1+\sqrt{-3}}{2}$ is a cube root of unity.

The integer

$$\Delta = \begin{cases} 4m & \text{if } m \equiv 2, 3, \bmod 4, \\ m & \text{if } m \equiv 1 \bmod 4, \end{cases}$$

is called the discriminant of $\mathcal{O}$.

The set K of all elements $a + b\sqrt{m}$ with $a, b \in \mathbb{Q}$ is called a quadratic number field. For $\alpha = a + b\sqrt{m}$ we call $\alpha' = a - b\sqrt{m}$ the conjugate of α, and $N\alpha = \alpha\alpha' = a^2 - mb^2 \in \mathbb{Q}$ the norm of α. Note that the norm is multiplicative: $N(\alpha\beta) = \alpha\beta\alpha'\beta' = N(\alpha)N(\beta)$.

The obvious connection with quadratic forms is the following: if $m \equiv 2, 3 \bmod 4$, then

$$X^2 - mY^2 = (X + Y\sqrt{m}\,)(X + Y\sqrt{m}\,) = N(X + Y\sqrt{m}\,),$$

and if $m \equiv 1 \bmod 4$, then

$$X^2 + XY + \frac{1-m}{4}Y^2 = N(X + Y\omega).$$

In fact, $\omega' = \frac{1-\sqrt{m}}{2} = 1 - \omega$, hence $N(X + Y\omega) = (X + Y\omega)(X + Y\omega') = X^2 + XY(\omega + \omega') + Y^2\omega\omega' = X^2 + XY + Y^2\frac{1-m}{4}$. Thus the principal quadratic form of discriminant Δ can be factored over K, and is in fact a norm.

As in $\mathbb{Z}$, we say that $\beta \mid \alpha$ in $\mathcal{O}$ if there is a $\gamma \in \mathcal{O}$ such that $\alpha = \beta\gamma$. Similarly, observe that $\mathcal{O} \setminus \{0\}$ is a monoid, so we automatically have the notion of divisibility, units, irreducibles and primes, and we know that all primes are irreducible.

In order to get familiar with these rings, let us prove some really simple results.

Lemma 5.8. *We have $n \mid a + b\omega$ for some $n \in \mathbb{Z}$ if and only if $n \mid a$ and $n \mid b$.*

Proof. We have $n \mid a + b\omega$ if and only if there is some $c + d\omega \in \mathcal{O}$ with $a + b\omega = n(c + d\omega) = nc + nd\omega$. Comparing coefficients shows that $a = nc$ and $b = nd$, and the converse is also clear. □

Next we determine the units.

Lemma 5.9. *An element $\alpha = a + b\omega \in \mathcal{O}$ is a unit if and only if $N\alpha = \pm 1$.*

Proof. Assume that α is a unit. Then $\alpha\beta = 1$ for some $\beta \in \mathcal{O}$, and taking the norms shows that $N\alpha N\beta = 1$, which implies $N\alpha = \pm 1$.

Conversely, if $N\alpha = \pm 1$, then $\alpha\alpha' = \pm 1$ and therefore $\pm\alpha\alpha' = 1$; but this means that α is a unit. □

In complex quadratic number fields, the norm is always nonnegative, hence $\alpha \in \mathcal{O}$ is a unit if and only if it has norm 1. In particular, $a + bi$ is a unit in $\mathbb{Z}[i]$ if and only if $a^2 + b^2 = 1$; thus the units in $\mathbb{Z}[i]$ are ± 1, $\pm i$. Similarly, the units in $Z[\omega]$ with discriminant $\Delta = -3$ are ± 1, $\pm\omega$, and $\pm\omega^2$. Note that $\omega^3 = -1$. If $\Delta < -3$, however, then $N\alpha = 1$ has only the trivial solutions $\alpha = \pm 1$.

If $\alpha = \beta\varepsilon$ for some unit $\varepsilon \in \mathcal{O}$, we say that α and β are associated, and write $\alpha \sim \beta$.

We can also introduce the notion of congruence by saying $\alpha \equiv \beta \bmod \mu$ if $\mu \mid (\alpha - \beta)$. Here, the same rules hold for $\mathcal{O}$ as for $\mathbb{Z}$: you can add, subtract, and multiply congruences. The following result will later be useful:

Lemma 5.10. *Let $\pi \in \mathcal{O}$ be an element such that $p = |N\pi|$ is prime. Then every $\alpha \equiv d \bmod \pi$ for some $d \in \{0, 1, \ldots, p-1\}$.*

Proof. Let $\pi = a + b\omega$. Note that $p \nmid b$ (otherwise $\pi \mid b$ in $\mathcal{O}$, hence $\pi \mid a$; taking norm gives $p \mid a$, hence $p \mid \pi$: contradiction). Thus there is some integer c with $bc \equiv 1 \bmod p$, and in particular we have $bc \equiv 1 \mod \pi$. Now $b\omega \equiv -a \bmod \pi$ implies $\omega \equiv bc\omega \equiv -ac \bmod \pi$. Thus given any $r + s\omega \in \mathcal{O}$, we have $r + s\omega \equiv r - sac \bmod \pi$, or in other words: every element in $\mathcal{O}$ is congruent to some rational integer modulo π. Since $r - sac \equiv d \bmod p$ for some $d \in \{0, 1, \ldots, p-1\}$, we also have $r + s\omega \equiv r - sac \equiv d \bmod \pi$. □

Actually the proof is valid whenever p is squarefree.

Lemma 5.11. *Every nonunit $\alpha \in \mathcal{O} \setminus \{0\}$ can be factored into irreducibles.*

Proof. We make induction on $|N\alpha|$. If $|N\alpha| = 1$, then α is a unit. Now assume that every nonunit α with $|N\alpha| < m$ has a factorization into irreducibles, and assume that $N\alpha = m$. If α is irreducible, we are done. If not, write $\alpha = \beta\gamma$ for nonunits β, γ. Since $m = |N\alpha| = |N\beta| \cdot |N\gamma|$, we see that $|N\beta|, |N\gamma| < m$. By induction assumption, $\beta = \pi_1 \cdots \pi_r$ and $\gamma = \rho_1 \cdots \rho_s$ have factorizations into irreducibles. But then $\alpha = \pi_1 \cdots \pi_r \cdot \rho_1 \cdots \rho_s$ is a factorization of α into irreducibles. □

We say that $\mathcal{O}$ has unique factorization if factorizations of α into irreducibles differ only by units.

Lemma 5.12. *The ring $\mathcal{O}$ has unique factorization if and only if every irreducible element in $\mathcal{O}$ is prime.*

Proof. Assume that $\alpha = \pi_1 \cdots \pi_r = \rho_1 \cdots \rho_s$ has two factorizations into irreducibles. Since π_1 is irreducible, it is prime, hence π_1 divides some ρ_j, say ρ_1. Since ρ_1 is irreducible, $\rho_1 = \pi_1\varepsilon$ for some unit. Cancelling π_1 and replacing ρ_2 by $\varepsilon\rho_2$ then shows $\pi_2 \cdots \pi_r = \rho_2 \cdots \rho_s$. After finitely many steps we will have proved that $r = s$ and $\pi_j \sim \rho_j$ for every j. □

In $\mathbb{Z}[i]$, the integer 3 is irreducible: if we had $3 = \alpha\beta$ for nonunits $\alpha, \beta \in \mathbb{Z}[i]$, then $9 = N\alpha N\beta$. Since α, β are not units, we must have $N\alpha = 3$, hence $3 = x^2 + y^2$ for $\alpha = x + yi$. But this equation is not solvable.

The integer 5, on the other hand, is not irreducible in $\mathbb{Z}[i]$ since $5 = (1 + 2i)(1 - 2i)$. The element $1 + 2i$ is irreducible (because its norm 5 is irreducible) and even prime (we will see this below).

Thus for proving that $\mathcal{O}$ is a unique factorization domain we have to show that irreducible elements are prime.

Not all rings $\mathbb{Z}[\omega]$, however, have unique factorization. Consider e.g. the field $\mathbb{Q}(\sqrt{-5}\,)$ and the ring $\mathbb{Z}[\sqrt{-5}\,]$ with discriminant $\Delta = -20$. Then

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}\,)(1 - \sqrt{-5}\,). \tag{5.1}$$

We claim that the elements 2, 3, $1 \pm \sqrt{-5}$ are all irreducible and do not differ by units (the last claim is obvious since the only units are ± 1). Once we have proved this, (5.1) will be an example of nonunique factorization in $\mathbb{Z}[\sqrt{-5}]$.

For showing that 2 is irreducible, assume $2 = \alpha\beta$ for nonunits $\alpha, \beta \in \mathbb{Z}[\sqrt{-5}]$. Taking the norm gives $4 = N\alpha N\beta$, which is an equation in integers. If $N\alpha = 1$, then α is a unit, hence we must have $N\alpha = N\beta = 2$. But the equation $x^2 + 5y^2 = 2$ is not solvable.

Lemma 5.13. *Let $p \in \mathbb{Z}$ be a prime. Then p is prime in $\mathcal{O}$ if and only if $(\frac{\Delta}{p}) = -1$.*

Proof. Assume first that $(\frac{\Delta}{p}) = -1$, and suppose that $p \mid \alpha\beta$ for $\alpha, \beta \in \mathcal{O}$. Taking norms gives $p^2 \mid N\alpha N\beta$. Since p is prime in $\mathbb{Z}$, it divides one of the factors, say $p \mid N\alpha$. But $N\alpha = N(x + y\omega) = Q(x, y)$ for the principal form Q, hence $p \mid \Delta$ or $(\frac{\Delta}{p}) = +1$ unless $p \mid x$ and $p \mid y$; but in this case, clearly $p \mid \alpha$.

Now assume that $p \mid \Delta$ or $(\frac{\Delta}{p}) = +1$. Leaving the case $p = 2$ as an exercise, we may assume that p is odd. Let us write $\Delta = 4m$ or $\Delta = m$ according as Δ is even or odd. Then $m \equiv b^2 \bmod p$ for some integer b, hence $p \mid b^2 - m = (b + \sqrt{m}\,)(b - \sqrt{m}\,)$, but clearly $p \nmid (b \pm \sqrt{m}\,)$. Thus p cannot be prime. □

In particular, the element 3 is prime in $\mathbb{Z}[i]$, and it is not prime in $\mathbb{Z}[\sqrt{-5}]$.

Lemma 5.14. *Let $\pi \in \mathcal{O}$ be an element such that $p = |N\pi|$ is prime. Then π is prime in $\mathcal{O}$.*

Proof. Assume that $\pi \mid \alpha\beta$, i.e., that $\alpha\beta \equiv 0 \bmod \pi$. By Lemma 5.10 there are integers $a, b \in \{0, 1, \ldots, p-1\}$ such that $\alpha \equiv a \bmod \pi$ and $\beta \equiv b \bmod p$. Thus $0 \equiv \alpha\beta \equiv ab \bmod \pi$, or $\pi \mid ab$. Taking norms shows that $p \mid ab$, and since p is prime, we have $p \mid a$ or $p \mid b$. Since $\pi \mid p$, the claim follows. □

Finally we need to be able to characterize irreducible elements:

Lemma 5.15. *Let $\mathcal{O}$ be a ring of integers with discriminant $\Delta < 0$, and assume that $h(\Delta) = 1$. If $\pi \in \mathcal{O}$ is irreducible, then either $\pi \sim p$ for some prime p with $(\frac{\Delta}{p}) = -1$, or $N\pi = p$ is prime.*

Proof. Assume that π is not associated to a rational prime. Then $\pi = a + b\omega$. If $d = \gcd(a, b) > 1$, let p be a prime dividing d. Then $p \mid \pi$, and since π is irreducible, we must have $\pi \sim p$ contradicting the assumption. Thus $\gcd(a, b) = 1$. But then $N\pi = Q(a, b)$ is properly represented by the principal form Q of discriminant Δ. Let p be a prime factor of $N\pi$; then p is also represented by some form of discriminant Δ, and since $h(\Delta) = 1$, it is represented by Q, i.e., we have $p = Q(c, d)$. Put $\rho = c + d\omega$; since ρ is prime and $\rho \mid p$ and $p \mid N\pi = \pi\pi'$, we conclude that $\rho \mid \pi$ or $\rho \mid pi'$; the latter relation is equivalent to $\rho' \mid \pi$. Since π is irreducible, we must have $\pi \sim \rho$ or $\pi \sim \rho'$. In both cases, we see that $|N\pi| = |N\rho| = p$. □

Now we can prove

Theorem 5.16. *Let $\mathcal{O}$ be a ring of integers with discriminant $\Delta < 0$. Then $h(\Delta) = 1$ if and only if $\mathcal{O}$ is a unique factorization domain.*

Proof. Assume first that $h(\Delta) = 1$. Let π be irreducible. Then either $\pi \sim p$ for some prime p with $(\frac{\Delta}{p}) = -1$ (and these π are prime by Lemma 5.13), or $|N\pi| = p$ is prime, and in this case π is prime by Lemma 5.15. Thus every irreducible is prime, and this implies unique factorization.

Now assume that $\mathcal{O}$ has unique factorization, and let p be a prime with $(\frac{\Delta}{p}) \neq -1$. Then p is not prime in $\mathcal{O}$, hence it is not irreducible. From $p = \pi\rho$ we get $N\pi = p$, hence p is represented by the principal form $Q = (1, 0, m)$ or $Q = (1, 1, m)$, according as $\Delta = -4m$ or $\Delta = 1 - 4m$.

Since $h(\Delta) = 1$ for $\Delta = -3, -4$, we may assume that $m > 1$. Now let (A, B, C) be a reduced form of discriminant Δ. If $A = 1$, then it is easily seen that $(A, B, C) = Q$. Assume therefore $A > 1$, and write $A = p_1 \cdots p_r$ for primes p_i. Since $(\Delta/p_i) = (B^2/p_i) \neq -1$, we have $p_i = N\pi_i$ for some $\pi_i \in \mathcal{O}$, hence $A = N\alpha$ for $\alpha = \pi_1 \cdots \pi_t$. Thus $A > 1$ is represented by Q; since the smallest integers represented by Q are 1 and m, we conclude that $A \geq m$. From $m \leq A \leq \sqrt{-\Delta/3}$ we deduce that $3m^2 \leq -\Delta$, which gives $3m^2 \leq 4m$ if $\Delta = -4m$, and $3m^2 \leq 4m - 1$ if $\Delta = 1 - 4m$. Both inequalities imply $m \leq 1$, and this contradiction proves that there is no reduced form with $A > 1$. □

Quadratic rings with unique factorization have been applied to the solution of diophantine equations already by Euler:

Proposition 5.17. *The only integral solutions of the diophantine equation $y^2 = x^3 - 2$ are $(x, y) = (3, \pm 5)$.*

Proof. We write $x^3 = y^2 + 2 = (y + \sqrt{-2}\,)(y - \sqrt{-2}\,)$. Observe that $\mathbb{Z}[\sqrt{-2}\,]$ has unique factorization, and that its only units are ± 1. Any gcd of the two factors on the right divides their difference $2\sqrt{-2}$, as well as their product x^3. But x is odd (if x is odd, then so is y, and then $x^3 = y^2 + 2 \equiv 2 \bmod 4$ cannot be a cube mod 4), hence not divisible by the prime $\sqrt{-2}$. Thus the factors $y \pm \sqrt{-2}$ are coprime, and their product is a cube. By unique factorization, each factor must be a unit times a cube, and since the units ± 1 are cubes themselves, we conclude that $y + \sqrt{-2} = (a + b\sqrt{-2}\,)^3$ for integers a, b. Multiplying out and comparing the imaginary parts gives $1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2)$. This implies $b = \pm 1$ and $a = \pm 1$, hence $y = a^3 - 6ab^2 = \pm 5$. But then $x = 3$, and this proves the claim. □

This approach (factoring a diophantine equation over a suitable number field) is very classical; Kummer and his successors tried to prove Fermat's Last Theorem in a similar way, but ultimately did not succeed (they could prove it for all prime exponents below $4 \cdot 10^9$): Wiles' proof uses completely different techniques.

Exercises

5.1 Show that $2 \cdot 5 = -\sqrt{-10} \cdot \sqrt{-10}$ is an example of nonunique factorization in $\mathbb{Z}[\sqrt{-10}\,]$.

5.2 Show that $2 \cdot 3 = -\sqrt{-6} \cdot \sqrt{-6}$ is an example of nonunique factorization in $\mathbb{Z}[\sqrt{-6}\,]$.

5.3 Generalize the two preceding exercises.

5.4 Show that there exist units $\neq \pm 1$ in $\mathbb{Z}[\sqrt{2}\,]$. Do the same for $\mathbb{Z}[\sqrt{3}\,]$ and $\mathbb{Z}[\sqrt{6}\,]$, as well as for $\mathbb{Z}[\frac{1+\sqrt{5}}{2}\,]$. It can be proved that the rings $\mathbb{Z}[\omega]$ with $\Delta > 0$ always have nontrivial units.

5.5 Show that $2 = \sqrt{2} \cdot \sqrt{2} = (2 + \sqrt{2}\,)(2 - \sqrt{2}\,)$ is not an example of nonunique factorization.

5.6 Show that $6 = 2 \cdot 3 = (3 + \sqrt{3}\,)(3 - \sqrt{3}\,)$ is not an example of nonunique factorization.

Author Index

Subject Index