MA257: INTRODUCTION TO NUMBER THEORY
LECTURE NOTES 2018

J. E. CREMONA

Contents

0. Introduction: What is Number Theory?

Number Theory is (of course) primarily the Theory of Numbers: ordinary whole numbers (integers). It is, arguably, the oldest branch of mathematics. Integer solutions to Pythagoras's equation

$$a^2 + b^2 = c^2$$

have been found, systematically listed with all the arithmetic carried out in base 60, on ancient Babylonian clay tablets. There are several different flavours of Number Theory, distinguished more by the methods used than by the problems whose solutions are sought. These are

- *Elementary* Number Theory: using elementary methods only;
- *Analytic* Number Theory: using analysis (real and complex), notably to study the distribution of primes;
- *Algebraic* Number Theory: using more advanced algebra, and also studying *algebraic numbers* such as $1 + \sqrt[3]{2} + \sqrt[17]{17}$;
- *Geometric* Number Theory: using geometric, algebraic and analytic methods; also known as *arithmetic algebraic geometry*.

Andrew Wiles used a vast array of new techniques and previously known results in arithmetic algebraic geometry to solve Fermat's Last Theorem, whose statement is entirely elementary (see below). This is typical of progress in Number Theory, where there have been many cases of entirely new mathematical theories being created to solve specific, and often quite elementary-seeming problems.

This module is mostly elementary with some analytic and algebraic parts. The algebraic approach is pursued further in the module MA3A6 (Algebraic Number Theory). The geometric approach is pursued further in the module MA426 (Elliptic Curves).

Number Theory starts out with simple questions about integers: simple to state, if not to answer. Here are three types of question:

- *Diophantine Equations* are equations to which one seeks integers solutions (or sometimes rational solutions). For example,
  (1) $x^2 + y^2 = z^2$ has infinitely many integral solutions (so-called Pythagorean triples); later, we will see how to find them all.
  (2) $x^n + y^n = z^n$ has *no* nonzero integer solutions when $n \geq 3$. This is Fermat's Last Theorem, which we will certainly not be proving in these lectures, though we will prove the case $n = 4$.
  (3) $y^2 = x^3 + 17$ has exactly 8 integer solutions (x, y), $x = -2, -1, 2, 4, 8, 43, 52$ and one further value which you can find for yourselves. Proving that there are no more solutions is harder; using Sage you can solve this as follows:

```
sage: EllipticCurve([0,17]).integral_points()
```

  (4) Every positive integer n can be written as a sum of four squares (including 0), for example

$$47 = 1 + 1 + 9 + 36 = 1^2 + 1^2 + 3^2 + 6^2,$$

  but not all may be written as a sum of 2 or 3 squares. Which?

```
sage: sum_of_k_squares(4,47)
```

  We will answer the two- and four-square problems later, with a partial answer for three squares.

- Questions about primes, for example

(1) There are infinitely many primes (an ancient result proved in Euclid.)

(2) Is every even number (greater than 2) expressible as the sum of two primes? This was conjectured by Goldbach in 1746 and still not proved, though it has been verified for numbers up to 4×10^{18}; the "weak form" of the conjecture, that every odd number greater than 5 is a sum of three primes, was proved in 2013 by the Peruvian Harald Helfgott.

(3) Are all the Fermat numbers $F_n = 2^{2^n} + 1$ prime (as Fermat also claimed)? Certainly not: the first four are ($F_1 = 5$, $F_2 = 17$, $F_3 = 257$, $F_4 = 65537$) but then $F_5 = 641 \times 6700417$, $F_6 = 274177 \times 67280421310721$, $F_7 = 59649589127497217 \times 5704689200685129054721$, and no more prime values have been discovered in the sequence.

```
sage: [(2^2^n+1).factor() for n in range(9)]
```

(4) How many primes end in the digit 7? Infinitely many? Of the 664579 primes less than 10 million, the number which end in the digits 1, 3, 7 and 9 respectively are 166104, 166230, 166211, and 166032 (that is, 24.99%, 25.01%, 25.01% and 24.98%). What does this suggest?

```
sage: pc=dict([(d,0) for d in range(10)])
sage: for p in prime_range(10^7): pc[p%10]+=1
sage: [(d,pc[d],100.0*pc[d]/sum(pc.values()))
          for d in [1,3,7,9]]
```

(5) Are there infinitely many so-called *prime pairs*: primes which differ by only 2, such as $(3, 5)$, $(71, 73)$ or $(1000000007, 1000000009)$?

- Efficient algorithms for basic arithmetic: many modern applications of Number Theory are in the field of cryptography (secure communication of secrets, such as transmitting confidential information over the Internet). These application rely on the fact that the following two questions, which seem trivial from the theoretical points of view, are not at all trivial when asked about very large numbers with dozens or hundreds of digits:
  (1) Primality Testing: given a positive integer n, determine whether n is prime;
  (2) Factorization: given a positive integer n, determine the prime factors of n.

In this module, we will study a variety of such problems, mainly of the first two types, while also laying the theoretical foundations to further study.

Basic Notation. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ will denote, as usual, the sets of integers, rational numbers, real numbers and complex numbers. The integers form a ring, the others sets are fields.

$\mathbb{N} = \{n \in \mathbb{Z} \mid n \geq 1\}$ is the set of *natural numbers* (positive integers).

$\mathbb{N}_0 = \{n \in \mathbb{Z} \mid n \geq 0\}$ is the set of non-negative integers.

$\mathbb{P}$ will denote the set of (positive) prime numbers: integers $p > 1$ which have no factorization $p = ab$ with $a, b > 1$.

Divisibility: for $a, b \in \mathbb{Z}$ we write $a|b$, and say a *divides* b, when b is a multiple of a:

$$a|b \iff \exists c \in \mathbb{Z} : b = ac.$$

If a does not divide b we write $a \nmid b$. The divisibility relation gives a partial order on $\mathbb{N}$ with 1 as the "least" element and no "greatest element".

Congruence: for $a, b, c \in \mathbb{Z}$ with $c \neq 0$ we write $a \equiv b \pmod{c}$ and say that a is congruent to b modulo c if $c|(a - b)$:

$$a \equiv b \pmod{c} \iff c|(a - b).$$

Divisibility and congruence will be studied in detail later.

1. Factorization

1.1. Divisibility in $\mathbb{Z}$.

Definition 1.1.1. *Let $a, b \in \mathbb{Z}$. Then we say that a divides b and write $a|b$ if $b = ac$ for some $c \in \mathbb{Z}$:*

$$a|b \iff \exists c \in \mathbb{Z} : b = ac.$$

Alternatively, we may say that "b is a multiple of a". If $a \neq 0$ this is equivalent to the statement that the rational number b/a is an integer c. If a does not divide b we write $a \nmid b$.

Lemma 1.1.2. *[Easy facts about divisibility] For all a, b, $\ldots \in \mathbb{Z}$:*

(1) $a|b \implies a|kb \quad (\forall k \in \mathbb{Z})$;
(2) $a|b_1$, $a|b_2 \implies a|b_1 \pm b_2$; *hence if b_1 and b_2 are multiples of a, then so are all integers of the form $k_1b_1 + k_2b_2$.*
(3) $a|b$, $b|c \implies a|c$;
(4) $a|b$, $b|a \iff a = \pm b$;
(5) $a|b$, $b \neq 0 \implies |a| \leq |b|$; *so nonzero integers have only a finite number of divisors;*
(6) *If $k \neq 0$ then $a|b \iff ka|kb$;*
(7) *Special properties of ± 1: $\pm 1|a \quad (\forall a \in \mathbb{Z})$, and $a| \pm 1 \iff a = \pm 1$;*
(8) *Special properties of 0: $a|0 \quad (\forall a \in \mathbb{Z})$, and $0|a \iff a = 0$.*

Proposition 1.1.3 (Division Algorithm in $\mathbb{Z}$). *Let $a, b \in \mathbb{Z}$ with $a \neq 0$. There exist unique integers q, r such that*

$$b = aq + r \qquad \text{with} \qquad 0 \leq r < |a|.$$

Notation: the set of all multiples of a fixed integer a is denoted (a) or $a\mathbb{Z}$:

$$(a) = a\mathbb{Z} = \{ka \mid k \in \mathbb{Z}\}.$$

Then we have $a|b \iff b \in (a) \iff (a) \supseteq (b)$: "to contain is to divide". From Lemma 1.1.2(4) we have $(a) = (b) \iff a = \pm b$.

An *ideal* in a commutative ring R is a subset I of R satisfying

(i) $0 \in I$;
(ii) $a, b \in I \implies a \pm b \in I$;
(iii) $a \in I$, $r \in R \implies ra \in I$.

Notation: $I \lhd R$. For example, the set of all multiples of a fixed element a of R is the *principal ideal* denoted (a) or aR. We say that a *generates* the principal ideal (a). The other generators of (a) are the *associates* of a: elements $b = ua$ where u is a unit of R.

Proposition 1.1.4. *Every ideal in $\mathbb{Z}$ is principal.*

Definition 1.1.5. *A* Principal Ideal Domain *or PID is a (nonzero) commutative ring R such that*

(i) $ab = 0 \iff a = 0$ *or* $b = 0$;
(ii) *every ideal of R is principal.*

So $\mathbb{Z}$ is a principal ideal domain. Every nonzero ideal of $\mathbb{Z}$ has a unique positive generator.

1.2. Greatest Common Divisors in $\mathbb{Z}$.

Theorem 1.2.1. *Let $a, b \in \mathbb{Z}$.*

(1) *There exists a unique integer d satisfying*
   (i) $d|a$ *and* $d|b$;

(ii) *if* $c|a$ *and* $c|b$ *then* $c|d$*;*
(iii) $d \geq 0$*.*
(2) *The integer* d *can be expressed in the form* $d = au + bv$ *with* $u, v \in \mathbb{Z}$*.*

Definition 1.2.2. *For* $a, b \in \mathbb{Z}$ *we define the Greatest Common Divisor (or GCD) of* a *and* b *to be the integer* d *with the properties given in the theorem. Notation:* $\gcd(a, b)$*, or sometimes just* (a, b)*. Integers* a *and* b *are said to be coprime (or relatively prime) if* $\gcd(a, b) = 1$*.*

So integers are coprime if they have no common factors other than ± 1. The identity $\gcd(a, b) = au + bv$ is sometimes called *Bezout's identity*.

Corollary 1.2.3. *[Basic Properties of* $\gcd$*] For all* $a, b, k, m \in \mathbb{Z}$*:*
(1) a *and* b *are coprime iff there exist* $u, v \in \mathbb{Z}$ *such that* $au + bv = 1$*;*
(2) $\gcd(a, b) = \gcd(b, a) = \gcd(|a|, |b|)$*;*
(3) $\gcd(ka, kb) = |k| \gcd(a, b)$*;*
(4) $\gcd(a, 0) = |a|$*;* $\gcd(a, 1) = 1$*;*
(5) $\gcd(a, b) = \gcd(a, b + ka)$ *for all* $k \in \mathbb{Z}$*;*
(6) *if* $\gcd(a, m) = \gcd(b, m) = 1$ *then* $\gcd(ab, m) = 1$*;*
(7) *if* $\gcd(a, b) = 1$ *then* $\gcd(a^k, b^l) = 1$ *for all* $k, l \in \mathbb{N}$*.*

Lemma 1.2.4. *[Euler's Lemma] If* $a|bc$ *and* $\gcd(a, b) = 1$ *then* $a|c$*.*

If a_1, a_2, $\ldots$, a_n is any finite sequence of integers then we similarly find that the ideal they generate, $I = (a_1, a_1, \ldots, a_n) = \{k_1a_1 + k_2a_2 + \cdots + k_na_n \mid k_1, k_2, \ldots, k_n \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z}$, hence $I = (d)$ for a unique $d \geq 0$, and we define $d = \gcd(a_1, a_2, \ldots, a_n)$. We say that

a_1, a_2, ..., a_n are *coprime* if $\gcd(a_1, a_2, \ldots, a_n) = 1$. This is weaker than the condition that $\gcd(a_i, a_j) = 1$ for all $i \neq j$: for example, $\gcd(6, 10, 15) = 1$ since $6 + 10 - 15 = 1$, but no pair of $6, 10, 15$ is coprime. When $\gcd(a_i, a_j) = 1$ for all $i \neq j$, we say that the a_i are *pairwise coprime*.

Our proofs have been non-constructive. A very important computational tool is the Euclidean Algorithm, which computes $d = \gcd(a, b)$ given a and $b \in \mathbb{Z}$, and its extended form which also computes the (non-unique) u, v such that $d = au + bv$.

1.3. The Euclidean Algorithm in $\mathbb{Z}$.

The Euclidean Algorithm is an efficient method of computing $\gcd(a, b)$ for any two integers a and b, without having to factorize them. It may also be used to compute the coefficients u and v in the identity $d = \gcd(a, b) = au + bv$.

The basic idea is this. We may assume $b > a > 0$ (see the Basic Properties above). Write $r = b - aq$ with $0 \leq r < a$; then $\gcd(a, b) = \gcd(r, a)$ and we have reduced the problem to a smaller one. After a finite number of steps we reach 0, and the last positive integer in the sequence $a, b, r, \ldots$ is the $\gcd$.

Example: $(963, 657) = (657, 963) = (306, 657) = (45, 306) = (36, 45) = (9, 36) = (0, 9) = 9$. Here we have used $963 - 657 = 306$, $657 - 2 \cdot 306 = 45$, $306 - 6 \cdot 45 = 36$, $45 - 36 = 9$.

To solve $9 = 963u + 657v$ we can back-substitute in these equations: $9 = 45 - 36 = 45 - (306 - 6 \cdot 45) = 7 \cdot 45 - 306 = 7 \cdot (657 - 2 \cdot 306) - 306 = 7 \cdot 657 - 15 \cdot 306 = 7 \cdot 657 - 15(963 - 657) = 22 \cdot 657 - 15 \cdot 963$, so $u = -15$ and $v = 22$.

There is a simpler way of keeping track of all these coefficients while reducing the amount which needs to be written down, using some auxiliary variables, which leads to the Euclidean algorithm. We give it in a form which keeps all the auxiliary variables positive which is easier to carry out in practice.

Extended Euclidean Algorithm: Given positive integers a and b, this algorithm computes (d, u, v) such that $d = \gcd(a, b) = au + bv$:

(1) Set $a_1 = a$, $a_2 = b$; $x_1 = 1$, $x_2 = 0$; $y_1 = 0$, $y_2 = 1$.
(2) Let $q = [a_1/a_2]$.
(3) Set $a_3 = a_1 - qa_2$; $x_3 = x_1 + qx_2$; $y_3 = y_1 + qy_2$.
(4) Set $a_1 = a_2$, $a_2 = a_3$; $x_1 = x_2$, $x_2 = x_3$; $y_1 = y_2$, $y_2 = y_3$.
(5) If $a_2 > 0$ loop back to Step 2.
(6) If $ax_1 - by_1 > 0$ return $(d, u, v) = (a_1, x_1, -y_1)$, else return $(d, u, v) = (a_1, -x_1, y_1)$.

Example: In the previous example, the a_i sequence is

$$963, 657, 306, 45, 36, 9, 0$$

using quotients

$$q = 1, 2, 6, 1, 4.$$

So the x_i sequence is

$$1, 0, 1, 2, 13, 15, 73$$

and the y_i sequence is

$$0, 1, 1, 3, 19, 22, 107.$$

Using the last x_i and y_i provides a check:

$$73a - 107b = 73 \cdot 963 - 107 \cdot 657 = 0$$

and the preceding values give the solution:

$$15a - 22b = 15 \cdot 963 - 22 \cdot 657 = -9.$$

So we may take $u = -15$, $v = 22$.

1.4. Primes and unique factorization.

Definition 1.4.1. *A prime number (or prime for short) is an integer $p > 1$ whose only divisors are ± 1 and $\pm p$; the set of primes is denoted $\mathbb{P}$:*

$$p \in \mathbb{P} \iff p > 1 \quad \text{and} \quad p = ab \implies a = \pm 1 \quad \text{or} \quad b = \pm 1.$$

For example $2, 3, 5, 7, 11$ are primes. Integers $n > 1$ which are not prime are called *composite*. If a is any integer then either $p|a$, in which case $\gcd(p, a) = p$, or $p \nmid a$, in which case $\gcd(p, a) = 1$.

Lemma 1.4.2. *Let p be a prime and $a, b \in \mathbb{Z}$. If $p|ab$ then either $p|a$ or $p|b$ (or both).*

This property of primes is very important, and the uniqueness of prime factorization relies on it. (It is easy to see that composite numbers do not have this property.) More generally:

Corollary 1.4.3. *Let p be a prime and $a_1, a_2, \ldots, a_n \in \mathbb{Z}$. Then*

$$p|a_1 a_2 \ldots a_n \implies p|a_i \quad \text{for some } i.$$

Theorem 1.4.4 (Fundamental Theorem of Arithmetic)**.** *Every positive integer n is a product of prime numbers, and its factorization into primes is unique up to the order of the factors.*

Note that this includes $n = 1$ which is an “empty” product, and primes themselves with only one factor in the product. Collecting together any powers of primes which occur in a prime factorization, we obtain

Corollary 1.4.5. *Every positive integer n may be expressed uniquely in the form*

$$n = p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}$$

where $p_1, \dots, p_k$ are primes with $p_1 < p_2 < \cdots < p_k$ and each $e_i \geq 1$. Alternatively, every positive integer n may be expressed uniquely in the form

$$n = \prod_{p \in \mathbb{P}} p^{e_p}$$

where the product is over all *primes, each $e_p \geq 0$, but only a* finite *number of $e_p > 0$.*

The exponent e_p which appears in this standard factorization of n is denoted $\text{ord}_p(n)$; it is characterized by the following property:

$$e = \text{ord}_p(n) \iff p^e | n \quad \text{and} \quad p^{e+1} \nmid n.$$

For example, $700 = 2^2 \cdot 5^2 \cdot 7$, so $\text{ord}_2(700) = \text{ord}_5(700) = 2$, $\text{ord}_7(700) = 1$, and $\text{ord}_p(700) = 0$ for primes $p \neq 2, 5, 7$. Every positive integer n is uniquely determined by the sequence of exponents $\text{ord}_p(n)$.

This standard factorization of positive integers into primes may be extended to negative integers by allowing a factor ± 1 in front of the product, and to nonzero rational numbers by allowing the exponents to be negative. We may accordingly extend the function ord_p to $\mathbb{Q}^*$, by setting $\text{ord}_p(-n) = \text{ord}_p(n)$ and $\text{ord}_p(n/d) = \text{ord}_p(n) - \text{ord}_p(d)$ for nonzero rationals n/d. [You should check that this is well-defined, independent of the representation of the fraction n/d.] Then we have the following extension of the main theorem on unique factorization:

Corollary 1.4.6. *Every nonzero rational number x may be uniquely expressed in the form*

$$x = \pm \prod_{p \in \mathbb{P}} p^{\mathrm{ord}_p(x)}.$$

For example, $-72/91 = -2^3 3^2 7^{-1} 13^{-1}$.

Many facts about integers may easily be proved using their unique factorization into primes. For example:

Proposition 1.4.7. *Let $m, n \in \mathbb{Z}$ be nonzero. Then*

$$m = \pm n \iff \mathrm{ord}_p(m) = \mathrm{ord}_p(n) \quad \forall p \in \mathbb{P}.$$

The function ord_p works rather like a logarithm. The following is easy to check:

Proposition 1.4.8. *Let $m, n \in \mathbb{Z}$ be nonzero. Then $\mathrm{ord}_p(mn) = \mathrm{ord}_p(m) + \mathrm{ord}_p(n)$.*

The previous result looks elementary enough, but it is sufficient to imply the uniqueness of prime factorization: for if $n = \prod p^{e_p}$ is *any* factorization of n in to primes, applying ord_q to both sides (where q is some fixed prime) and using the Proposition gives

$$\mathrm{ord}_q(n) = \sum e_p \mathrm{ord}_q(p) = e_q,$$

since $\mathrm{ord}_q(q) = 1$ and $\mathrm{ord}_q(p) = 0$ when $p \neq q$. It follows that the exponents e_p are uniquely determined.

Proposition 1.4.9. *Let $n \in \mathbb{Z}$ be nonzero. Then n is a perfect square if and only if $n > 0$ and $\mathrm{ord}_p(n)$ is even for all primes p.*

We end this section with a famous and ancient result of Euclid.

Theorem 1.4.10. *[Euclid] The number of primes is infinite.*

Note that this proof actually shows how to construct a "new" prime from any given finite set of known primes. Variations of this proof can show that there are infinitely many primes of various special forms: see the Exercises.

1.5. **Unique Factorization Domains.** Theorem 1.4.4 (extended to include negative integers) may be expressed succinctly by the statement that $\mathbb{Z}$ is a *Unique Factorization Domain* or UFD. Roughly speaking, a UFD is a ring in which every element has an essentially unique factorization as a unit times a product of "prime" elements. Every PID is a UFD (but not conversely: $\mathbb{Z}[X]$ is a UFD but not a PID), and an important source of PIDs is rings which have a "division algorithm" similar to the one for $\mathbb{Z}$. Such rings are called Euclidean Domains, and we start by defining these.

Definition 1.5.1. (a) *A nonzero ring R is an Integral Domain if, for $a, b \in R$,*

$$ab = 0 \iff (a = 0 \quad \text{or} \quad b = 0).$$

(b) *A nonzero ring R is a Euclidean Domain or ED if it is an integral domain equipped with a function $\lambda : R - \{0\} \to \mathbb{N}_0$ such that, for $a, b \in R$ with $a \neq 0$, there exist $q, r \in R$ such that*

$$b = aq + r \qquad \text{with either } r = 0 \text{ or } \lambda(r) < \lambda(a).$$

Examples:

- $\mathbb{Z}$ is an ED with $\lambda(n) = |n|$: this is what Proposition 1.1.3 states (though note that the definition of an ED does not require q and r to be unique).
- Any field F is an ED with $\lambda(x) = 0$ for all $x \neq 0$; this is a degenerate example since we may always take $r = 0$ in division.
- If F is a field then the polynomial ring $F[X]$ is an ED, using the degree function $\lambda(f(X)) = \deg(f(X))$. The required division property is well-known, being just the usual long division for polynomials.

  It is important that F is a field here: for example, $\mathbb{Z}[X]$ is *not* Euclidean (exercise).
- The ring $\mathbb{Z}[i]$ of *Gaussian Integers* is defined as

$$\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\};$$

  it is a subring of $\mathbb{C}$. We will study this in some detail as it gives another example of a Euclidean Domain which is of interest in number theory, both for its own sake and also for proving some properties of the ordinary or "rational" integers $\mathbb{Z}$. The Euclidean function λ on $\mathbb{Z}[i]$ is usually called the *norm* and denoted N:

$$N(\alpha) = \alpha\overline{\alpha} = a^2 + b^2 \qquad \text{for } \alpha = a + bi \in \mathbb{Z}[i].$$

Theorem 1.5.2. *The ring $\mathbb{Z}[i]$ of Gaussian Integers is a Euclidean Domain.*

Lemma 1.5.3. *The norm function N on $\mathbb{Z}[i]$ has the following properties:*

(1) *Multiplicativity: for all α, $\beta \in \mathbb{Z}[i]$, $N(\alpha\beta) = N(\alpha)N(\beta)$;*

(2) *Positivity:* $N(0) = 0$, $N(\alpha) \geq 1$ *for* $\alpha \neq 0$;
(3) *Units:* $N(\alpha) = 1 \iff \alpha \in U(\mathbb{Z}[i]) = \{\pm 1, \pm i\}$.

Recall that for a ring R, the group of *units* (invertible elements) is denoted $U(R)$. Elements of an integral domain are called *associate* if one is a unit times the other, or (equivalently) if each divides the other.

Example: Take $\alpha = 3 + 4i$ and $\beta = 10 + 11i$. Then

$$\frac{10+11i}{3+4i} = \frac{(10+11i)(3-4i)}{25} = \frac{74-7i}{25} = 3 + \frac{-1-7i}{25},$$

so the quotient is 3 and remainder $(10 + 11i) - 3(3 + 4i) = 1 - i$. Check: $N(1 - i) = 2$ is less than $N(3 + 4i) = 25$.

Just as we did for $\mathbb{Z}$, we can now prove that every ED is a PID:

Theorem 1.5.4. *Let R be a Euclidean Domain. Then R is a Principal Ideal Domain.*

In a PID we have $\gcd$s just as in $\mathbb{Z}$, and Bezout's identity. In general we do not have uniqueness of $\gcd$s, only uniqueness up to associates (multiplication by a unit). (In $\mathbb{Z}$ we avoided this non-uniqueness by insisting that all $\gcd$s were non-negative.)

Definition 1.5.5. *In a ring R, a $\gcd$ of two elements a and b is an element d satisfying*
(i) $d|a$ *and* $d|b$;
(ii) *if* $c|a$ *and* $c|b$ *then* $c|d$.

Lemma 1.5.6. *If $\gcd(a, b)$ exists then it is unique up to associates.*

Because of this non-uniqueness we cannot talk about *the* $\gcd$, only *a* $\gcd$ of a and b. In specific rings, one may impose an extra condition to ensure uniqueness: in $\mathbb{Z}$ we insisted that $\gcd(a, b) \geq 0$; in the polynomial ring $F[X]$ (with F a field) one usually insists that $\gcd(a(X), b(X))$ is *monic* (with leading coefficient 1).

Proposition 1.5.7. *In a PID, the* $\gcd$ *of two elements* a *and* b *exists, and may be expressed in the form* $au + bv$.

So in a PID, whether Euclidean or not, the $\gcd$ always exists. However, it is only in a ED that computing $\gcd$s is easily possible via the Euclidean Algorithm.

Example: Take $\alpha = 3 + 4i$ and $\beta = 10 + 11i$. Then from the previous example we have $\beta - 3\alpha = 1 - i$. Similarly, $\alpha - 3i(1-i) = i$, and lastly $1 - i = i(-1-i)$ with zero remainder. The last nonzero remainder was i which is therefore a $\gcd$ of α and β; one would normally adjust this since i is a unit and say that $\gcd(\alpha, \beta) = 1$. Back-substitution gives $i = \alpha - 3i(\beta - 3\alpha) = (1+9i)\alpha - 3i\beta$, so finally $1 = (9 - i)\alpha - 3\beta$.

The next step is to show that every PID is also a unique factorization domain. In the case of $\mathbb{Z}$, we used the Euclidean property again, and not just the PID property, for this step, but since there are rings which are PIDs but not Euclidean we give a proof which works for all PIDs.

Definition 1.5.8. *In an integral domain* R*, an element* p *is called* irreducible *if it is neither* 0 *nor a unit and* $p = ab$ *implies that either* a *or* b *is a unit;* p *is called* prime *if it is neither* 0 *nor a unit and* $p|ab$ *implies that either* $p|a$ *or* $p|b$.

Lemma 1.5.9. *Every prime is irreducible. In a PID, every irreducible is prime.*

The last property will be crucial in proving the uniqueness of factorizations into irreducibles, but for the existence we need to do some more preparation. The following lemma is called the "ascending chain condition" or ACC for ideals in a PID.

Lemma 1.5.10. *Let R be a PID. Let $(a_i)_{i\in\mathbb{N}}$ be a sequence of elements of R with $(a_1) \subseteq (a_2) \subseteq (a_3) \subseteq \dots$. (So each a_i is a multiple of the next). Then there exists k such that $(a_k) = (a_{k+1}) = (a_{k+2}) = \dots$, so the chain of ideals stabilizes. Hence any strictly ascending chain of ideals $(a_1) \subset (a_2) \subset (a_3) \subset \dots$ must be finite.*

This lemma is used to replace induction in the proof of the existence of factorizations into irreducibles, which was used for $\mathbb{Z}$.

Proposition 1.5.11. *Let R be a PID. Every element of R which is neither 0 nor a unit is a product of irreducibles.*

Finally, we use the fact that in a PID irreducibles are prime to prove that the factorizations of any given nonzero non-unit are essentially the same, up to reordering the factors and replacing irreducibles by associates.

Definition 1.5.12. *An Integral Domain R is a* Unique Factorization Domain *or UFD if*

(i) *every nonzero element may be expressed as a unit times a product of irreducibles;*

(ii) *the factorization in (i) is unique up to the order of the factors and replacing the irreducibles by associates; that is, if $a \in R$ is nonzero and*

$$a = up_1p_2\dots p_r = vq_1q_2\dots q_s$$

with u, v units and all p_i, q_j irreducibles, then $r = s$, and after permuting the q_j if necessary, there are units v_j for $1 \le j \le r$ such that $q_j = v_j p_j$ and $u = v v_1 v_2 \dots v_r$.

Theorem 1.5.13. *Let R be a PID. Then R is a UFD.*

Example (continued): Since the ring $\mathbb{Z}[i]$ of Gaussian Integers is Euclidean, it is a PID and a UFD. We have already determined that its units are the four elements ± 1 and $\pm i$, but what are its primes/irreducibles?

(1) If $\pi \in \mathbb{Z}[i]$ is prime then π divides some ordinary "rational" prime p, since if $n = N(\pi) = \pi\overline{\pi}$ then $\pi | n$ so by primality of π, π divides at least one prime factor p of n.

(2) If $N(\pi) = p$ is prime, then π is irreducible: for if $\pi = \alpha\beta$ then $p = N(\pi) = N(\alpha)N(\beta)$, so one of α, β has norm 1 and is a unit. For example, $1 + i$, $2 + i$, $3 + 2i$, $4 + i$ are prime since their norms are 2, 5, 13, 17.

(3) If a rational prime p is a sum of two squares, $p = a^2 + b^2$, then setting $\pi = a + bi$ gives $p = N(\pi) = N(\overline{\pi})$, so π and $\overline{\pi}$ are both Gaussian primes. We will prove later, in Theorem 2.4.2, that every rational prime p of the form $4k + 1$ can be expressed in this way; the factors π and $\overline{\pi}$ are not associate (exercise).

(4) However, rational primes q of the form $4k + 3$ can *not* be expressed as sums of two squares, since squares all leave remainder of 0 or 1 when divided by 4, so all numbers of the form $a^2 + b^2$ leave a remainder of 0, 1 or 2 on division by 4. Such primes q remain prime in $\mathbb{Z}[i]$. For if $q = \alpha\beta$ with neither α nor β a unit, then $q^2 = N(\alpha)N(\beta)$ with both $N(\alpha)$, $N(\beta) > 1$, so (by unique factorization in $\mathbb{Z}$) we must have $N(\alpha) = N(\beta) = q$, so q would be a sum of two squares.

We sum up this example as follows; we have proved everything stated here except for the fact that all primes of the form $4k+1$ are sums of two squares (Theorem 2.4.2), and the remark about associates (exercise).

Theorem 1.5.14. *The ring $\mathbb{Z}[i]$ of Gaussian Integers is a Euclidean Domain and hence also a Principal Ideal Domain and a Unique Factorization Domain. Its units are the four elements ± 1, $\pm i$. Its primes are as follows (together with their associates):*

(1) $1+i$, *of norm* 2;
(2) *each rational prime p of the form $4k+1$ is a sum of two squares, $p = a^2+b^2$, and p factorizes in $\mathbb{Z}[i]$ as $p = \pi\overline{\pi}$ where $\pi = a+bi$ and $\overline{\pi} = a-bi$ are non-associate Gaussian primes of norm p;*
(3) *each rational prime q of the form $4k+3$ is also a Gaussian prime.*

For example, here are some Gaussian factorizations: $123+456i = 3\cdot(1+2i)\cdot(69+14i)$ (the last factor has prime norm 4957), $2000 = (1+i)^8(1+2i)^3(1-2i)^3$.

```
sage: Qi.<i> = QQ.extension(x^2+1)
sage: 2018.factor()
2 * 1009
sage: Qi(2018).factor()
(i) * (15*i - 28) * (i + 1)^2 * (15*i + 28)
sage: (123+456*i).norm().factor()
3^2 * 5 * 4957
sage: (123+456*i).factor()
(-1) * (-14*i - 69) * (2*i + 1) * 3
```

There are other "number rings" similar to $\mathbb{Z}[i]$, but not many which are known to have unique factorization. A complete study requires more algebra, and is done in Algebraic Number Theory. Here are some further examples.

Example: The ring $R = \mathbb{Z}[\sqrt{-2}]$ is also Euclidean and hence a UFD. The proof is almost identical to the one given above for $\mathbb{Z}[i]$, using the norm $N(\alpha) = \alpha\overline{\alpha}$, so that $N(a + b\sqrt{-2}) = a^2 + 2b^2$. The key fact which makes R Euclidean via the norm is that every point in the complex plane is at distance less than 1 from the nearest element of R, as was the case with $\mathbb{Z}[i]$. Factorization of primes p now depends on $p \pmod 8$.

Example: The ring $R = \mathbb{Z}[\sqrt{-3}]$ is **not** Euclidean, and neither a PID nor a UFD. For example, $4 = 2 \cdot 2 = (1 + \sqrt{-3}) \cdot (1 - \sqrt{-3})$ with all factors on the right irreducible in R. Also: the ideal $(2, 1 + \sqrt{-3})$ is not principal; and the element 2 is irreducible but not prime (as the previous equation shows, since neither $1 \pm \sqrt{-3}$ are divisible by 2 in R). However, if we enlarge the ring by including numbers of the form $(a+b\sqrt{-3})/2$ where a and b are both odd, we obtain the larger ring $S = \mathbb{Z}[\omega]$, where $\omega = (-1 + \sqrt{-3})/2$, satisfying $\omega^2 + \omega + 1 = 0$, which is Euclidean and hence a UFD. The norm is again $N(\alpha) = \alpha\overline{\alpha}$; with $\alpha = a + b\omega$ one computes that $N(\alpha) = a^2 - ab + b^2$, and $4N(\alpha) = (2a - b)^2 + 3b^2$. This ring turns out to be useful in the solution of the Fermat equation $x^3 + y^3 = z^3$.

Example: As in the previous example, the ring $\mathbb{Z}[\sqrt{-19}]$ is not Euclidean. Enlarging it to $R = \mathbb{Z}[\omega]$, where now $\omega = (-1 + \sqrt{-19})/2$, satisfying $w^2 + w + 5 = 0$, we find a ring which is still not Euclidean, but is a PID and hence a UFD. This example shows that not every PID is Euclidean. We omit the details.

2. Congruences and modular arithmetic

The notation for congruence is an invention of Gauss. It simplifies many calculations and arguments in number theory.

2.1. Definition and Basic Properties.

Definition 2.1.1. *Let m be a positive integer. For $a, b \in \mathbb{Z}$ we say that a is congruent to b modulo m and write $a \equiv b \pmod{m}$ iff $a - b$ is a multiple of m:*

$$a \equiv b \pmod{m} \iff m|(a-b).$$

Here m is called the modulus. If $m \not| (a-b)$ then we write $a \not\equiv b \pmod{m}$.

For example, $-3 \equiv 18 \pmod{7}$ and $19 \not\equiv 1 \pmod{4}$. All even integers are congruent to $0 \pmod{2}$, while odd integers are congruent to $1 \pmod{2}$.

Congruence may be expressed in algebraic terms: to say $a \equiv b \pmod{m}$ is equivalent to saying that the cosets $a + m\mathbb{Z}$ and $b + m\mathbb{Z}$ of $m\mathbb{Z}$ in $\mathbb{Z}$ are equal.

The basic properties of congruence are summarized in the following lemmas.

Lemma 2.1.2. *For each fixed modulus m, congruence modulo m is an equivalence relation:*
(i) *Reflexive: $a \equiv a \pmod{m}$ for all $a \in \mathbb{Z}$;*
(ii) *Symmetric: $a \equiv b \pmod{m} \implies b \equiv a \pmod{m}$;*
(iii) *Transitive: If $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$ then $a \equiv c \pmod{m}$.*

Lemma 2.1.3. *If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$ then $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$.*

The preceding result has the following interpretation. As well as $m\mathbb{Z}$ being a subgroup of the additive group $\mathbb{Z}$, it is also an ideal of the ring $\mathbb{Z}$, and hence there is a well-defined quotient ring $\mathbb{Z}/m\mathbb{Z}$. The lemma says that addition and multiplication in $\mathbb{Z}/m\mathbb{Z}$ are well-defined. We will return to this viewpoint in the next section.

Lemma 2.1.4. (i) *If* $a \equiv b \pmod{m}$ *then* $ac \equiv bc \pmod{mc}$ *for all* $c > 0$*;*
(ii) *If* $a \equiv b \pmod{m}$ *and* $n|m$ *then* $a \equiv b \pmod{n}$*.*

Lemma 2.1.5. *If* $ax \equiv ay \pmod{m}$*, then* $x \equiv y \pmod{m/\gcd(a,m)}$*.*
Two important special cases:
If $ax \equiv ay \pmod{m}$ *and* $\gcd(a,m) = 1$*, then* $x \equiv y \pmod{m}$*.*
If $ax \equiv ay \pmod{m}$ *and* $a|m$*, then* $x \equiv y \pmod{m/a}$*.*

Proposition 2.1.6. *Let* $a, b \in \mathbb{Z}$*. The congruence* $ax \equiv b \pmod{m}$ *has a solution* $x \in \mathbb{Z}$ *if and only if* $\gcd(a,m)|b$*. If a solution exists it is unique modulo* $m/\gcd(a,m)$*.*
In particular, when $\gcd(a,m) = 1$ *the congruence* $ax \equiv b \pmod{m}$ *has a solution for every* b*, which is unique modulo* m*.*

How to solve the congruence $ax \equiv b \pmod{m}$: Use the EEA to find d, u, v with $d = \gcd(a,m) = au + mv$. Check that $d|b$ (otherwise there are no solutions). If $b = dc$ then $b = auc + mvc$ so $x = uc$ is one solution. The general solution is $x = uc + tm/d = (ub + tm)/d$ for arbitrary $t \in \mathbb{Z}$.

Lemma 2.1.7. *Each integer* a *is congruent modulo* m *to exactly one integer in the set* $\{0, 1, 2, \ldots, m-1\}$*. More generally, let* k *be a fixed integer. Then every integer is congruent modulo* m *to exactly one integer in the set* $\{k, k+1, k+2, \ldots, k+m-1\}$*.*

Definition 2.1.8. *Taking $k = 0$, we obtain the system of* least non-negative residues modulo m*:* $\{0, 1, 2, \dots, m-1\}$*. Taking $k = -[(m-1)/2]$ gives the system of* least residues modulo m*; when m is odd this is $\{0, \pm 1, \pm 2, \dots, \pm(m-1)/2\}$, while when m is even we include $m/2$ but not $-m/2$. Any set of m integers representing all m residue classes modulo m is called a* residue system modulo m*.*

For example, when $m = 7$ the least non-negative residues are $\{0, 1, 2, 3, 4, 5, 6\}$ and the least residues are $\{-3, -2, -1, 0, 1, 2, 3\}$; for $m = 8$ we have least non-negative residues $\{0, 1, 2, 3, 4, 5, 6, 7$ and least residues $\{-3, -2, -1, 0, 1, 2, 3, 4\}$.

2.2. **The structure of $\mathbb{Z}/m\mathbb{Z}$.**

Definition 2.2.1. *The* ring of integers modulo m *is the quotient ring $\mathbb{Z}/m\mathbb{Z}$. We will denote the group of units of $\mathbb{Z}/m\mathbb{Z}$ by U_m, and its order by $\varphi(m)$. The function $\varphi : \mathbb{N} \to \mathbb{N}$ is called* Euler's totient function *or* Euler's phi function*.*

Sometimes $\mathbb{Z}/m\mathbb{Z}$ is denoted $\mathbb{Z}_m$; however there is a conflict of notation here, since for prime p the notation $\mathbb{Z}_p$ is used to denote a different ring important in number theory, the ring of p-adic integers. *We will therefore not use this abbreviation!*

Informally we may identify $\mathbb{Z}/m\mathbb{Z}$ with the set $\{0, 1, 2, \dots, m-1\}$, though the elements of $\mathbb{Z}/m\mathbb{Z}$ are not integers but "integers modulo m": elements of the quotient ring $\mathbb{Z}/m\mathbb{Z}$. To be strictly correct, one should use the notation a, b, ... for integers and $\overline{a}$, $\overline{b}$, ... for their residues in $\mathbb{Z}/m\mathbb{Z}$. Then one has $\overline{a} = \overline{b}$ (in $\mathbb{Z}/m\mathbb{Z}$) iff $a \equiv b \pmod{m}$ (in $\mathbb{Z}$), and $\mathbb{Z}/m\mathbb{Z} = \{\overline{0}, \overline{1}, \overline{2}, \dots, \overline{m-1}\}$. For simplicity we will not do this but use the same notation for an integer and its residue in $\mathbb{Z}/m\mathbb{Z}$.

So $\mathbb{Z}/m\mathbb{Z}$ is a finite ring with m elements, and its unit group U_m is a finite group under the operation of "multiplication modulo m".

Proposition 2.2.2. *Let $a \in \mathbb{Z}/m\mathbb{Z}$. Then $a \in U_m$ (that is, a is invertible modulo m) if and only if $\gcd(a, m) = 1$.*

Remark: Note that if $a \equiv a' \pmod{m}$ then $\gcd(a, m) = \gcd(a', m)$, since $a' = a + km$ for some k. Hence the quantity $\gcd(a, m)$ only depends on the residue of a modulo m.

We may use the Extended Euclidean Algorithm to detect whether or not a is invertible modulo m, and also to find its inverse a' if so, since if (x, y) is a solution to $ax + my = 1$ then $ax \equiv 1 \pmod{m}$ so we may take $a' = x$. For example, $\gcd(4, 13) = 1$ with $4 \cdot 10 - 13 \cdot 3 = 1$, so the inverse of 4 modulo 13 is 10. Here is a complete table of inverses modulo 13:

a	0	1	2	3	4	5	6	7	8	9	10	11	12
a'	-	1	7	9	10	8	11	2	5	3	4	6	12

It follows that $\varphi(m)$, the order of U_m, is equal to the number of residues modulo m of integers which are coprime to m. This is often given as the definition of $\varphi(m)$.

Corollary 2.2.3.

$$\varphi(m) = |\{a \mid 0 \leq a \leq m-1 \quad \textit{and} \quad \gcd(a, m) = 1\}|.$$

Definition 2.2.4. *A* reduced residue system modulo m *is a set of $\varphi(m)$ integers covering the residue classes in U_m.*

Any set of $\varphi(m)$ integers which are all coprime to m, and no two of which are congruent modulo m, form a reduced residue system. The "standard" one is

$$\{a \mid 0 \leq a \leq m-1 \quad \text{and} \quad \gcd(a,m)=1\}.$$

For example, $U_6 = \{1,5\}$, $U_7 = \{1,2,3,4,5,6\}$ and $U_8 = \{1,3,5,7\}$, so that $\varphi(6)=2$, $\varphi(7)=6$ and $\varphi(8)=4$. Here are the first few values of $\varphi(m)$:

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1
$\varphi(m)$	1	1	2	2	4	2	6	4	6	4	10	4	12	6	

Proposition 2.2.5. (1) *$\varphi(m)$ is even for $m \geq 3$;*
(2) *$\varphi(m) = m-1$ if and only if m is prime;*
(3) *Let p be a prime; then $\varphi(p^e) = p^{e-1}(p-1)$ for $e \geq 1$.*

We will use this to obtain a general formula for $\varphi(m)$ after the Chinese Remainder Theorem below, which will reduce the determination of $\varphi(m)$ for general m to the case of prime powers.

Arithmetic modulo m is much simpler when m is prime, as the following result indicates.

Theorem 2.2.6. *If p is a prime then $\mathbb{Z}/p\mathbb{Z}$ is a field. If m is composite then $\mathbb{Z}/m\mathbb{Z}$ is not a field, and not even an integral domain.*

Notation: To emphasize its field structure, $\mathbb{Z}/p\mathbb{Z}$ is also denoted $\mathbb{F}_p$, and the multiplicative group U_p is then denoted $\mathbb{F}_p^*$. It has order $p-1$, and is cyclic (see Theorem 2.6.1 below).

2.3. **Euler's, Fermat's and Wilson's Theorems.** Since U_m is a finite multiplicative group of order $\varphi(m)$ we immediately have the following as a consequence of Lagrange's Theorem for finite groups.

Theorem 2.3.1. (a) **Euler's Theorem:** *Let m be a positive integer and a an integer coprime to m. Then*

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

(b) **Fermat's Little Theorem:** *Let p be a prime and a an integer not divisible by p. Then*

$$a^{p-1} \equiv 1 \pmod{p};$$

moreover, for every integer a we have

$$a^p \equiv a \pmod{p}.$$

Fermat's Little Theorem can be used as a primality test. Let n be an odd integer which one suspects to be a prime; if $2^{n-1} \not\equiv 1 \pmod{n}$ then n is certainly not prime. Note that this has been proved without exhibiting a factorization of n. On the other hand, if $2^{n-1} \equiv 1 \pmod{n}$ it does not prove that n is prime! For example this holds with $n = 1729 = 7 \cdot 13 \cdot 19$. Such a number is called a pseudoprime to base 2. By using a combination of so-called bases (as here we used the base 2) one can develop much stronger "probabilistic primality tests".

Corollary 2.3.2. *In $\mathbb{F}_p[X]$ the polynomial $X^p - X$ factorizes as a product of p linear factors:*

$$X^p - X = \prod_{a \in \mathbb{F}_p} (X - a) \qquad \text{in } \mathbb{F}_p[X].$$

Corollary 2.3.3. *[Wilson's Theorem] Let p be a prime. Then*

$$(p-1)! \equiv -1 \pmod{p}.$$

Remark: The converse to Wilson's Theorem also holds; in fact, for composite integers m greater than 4 we have $(m-1)! \equiv 0 \pmod{m}$ (exercise). But this is not useful as a primality test, since there is no way to compute the residue of $(m-1)! \pmod{m}$ quickly.
Example: Take $p = 13$. Then $(p-1)! = 12! = 479001600 = 13 \cdot 36846277 - 1$. A better way of seeing this is to write

$$12! \equiv 1 \cdot 12 \cdot (2 \cdot 7) \cdot (3 \cdot 9) \cdot (4 \cdot 10) \cdot (5 \cdot 8) \cdot (6 \cdot 11) \equiv 12 \equiv -1 \pmod{13}.$$

A similar trick, pairing each residue apart from ± 1 with its inverse, may be used to prove Wilson's Theorem directly. This works because ± 1 are the only residues modulo a prime which are their own inverse:

Proposition 2.3.4. *Let p be a prime. Then the only solutions to $x^2 \equiv 1 \pmod{p}$ are $x \equiv \pm 1$.*

Example: Let $m = F_5 = 2^{32} + 1 = 4294967297$. Check that $x = 1366885067$ satisfies $x^2 \equiv 1 \pmod{m}$. This proves that m is not prime. In fact, $m = ab$ where $a = 671 = \gcd(m, x-1)$ and $b = 6700417 = \gcd(m, x+1)$. Many modern factorization methods are based on this idea. Of course, one needs efficient ways to find solutions other than ± 1 to the congruence $x^2 \equiv 1 \pmod{m}$ where m is the (odd) composite number being factorized. There are several of these, which collectively go by the name of "quadratic sieve" methods.

2.4. Some Applications.

Proposition 2.4.1. *Let p be an odd prime. Then the congruence $x^2 \equiv -1 \pmod{p}$ has a solution if and only if $p \equiv 1 \pmod{4}$.*

There are many other ways of proving the preceding Proposition. One is to use the fact that $\mathbb{F}_p^*$ is cyclic (Theorem 2.6.1), hence has elements of order d for all $d|(p-1)$, and an element a of order 4 satisfies $a^4 = 1$, $a^2 \neq 1$, so $a^2 = -1$. Alternatively, from Wilson's Theorem one can show that for all odd p,

$$(((p-1)/2)!)^2 \equiv -(-1)^{(p-1)/2} \pmod{p},$$

so when $p \equiv 1 \pmod{4}$ the number $a = ((p-1)/2)!$ satisfies $a^2 \equiv -1 \pmod{p}$.

As a corollary we can prove the result used earlier, that a prime of the form $4k+1$ may be written as a sum of two squares.

Theorem 2.4.2. *Let p be a prime such that $p \equiv 1 \pmod{4}$. Then there exist integers a and b such that $p = a^2 + b^2$.*

Remarks The first proof can be made constructive: given c satisfying $c^2 \equiv -1 \pmod{p}$, it is not hard to show that the element $a + bi = \gcd(c + i, p)$ in $\mathbb{Z}[i]$ satisfies $a^2 + b^2 = p$, so a single application of the Euclidean algorithm in $\mathbb{Z}[i]$ gives a solution.

The first proof also shows that the solution is essentially unique, up to permuting a and b and changing their signs. This follows from the fact that the factorization of p in $\mathbb{Z}[i]$ as $p = \pi\overline{\pi}$ with $\pi = a + bi$ is unique up to permuting the factors and multiplying them by units.

We finish this section with some more applications to the distribution of primes.

Proposition 2.4.3. (a) *There are infinitely many primes* $p \equiv 1 \pmod 4$.
(b) *There are infinitely many primes* $p \equiv 3 \pmod 4$.

Similarly, odd prime divisors of $n^4 + 1$ are $\equiv 1 \pmod 8$ and there are therefore infinitely many of those; odd prime divisors of n^8+1 are $\equiv 1 \pmod{16}$ so there are infinitely many of those; and so on. Next we have

Proposition 2.4.4. *Let* q *be an odd prime.*
(a) *Let* p *be a prime divisor of* $f(n) = n^{q-1} + n^{q-2} + \cdots + n + 1$. *Then either* $p = q$ *or* $p \equiv 1 \pmod q$.
(b) *There are infinitely many primes* $p \equiv 1 \pmod q$.

Using *cyclotomic polynomials* (for example, $f(n)$ above) one can show that there are infinitely many primes $p \equiv 1 \pmod m$ for any m. More generally *Dirichlet's Theorem on primes in arithmetic progressions* states that there are infinitely many primes $p \equiv a \pmod m$ whenever a and m are coprime: the general proof uses complex analysis!

2.5. **The Chinese Remainder Theorem or CRT.**

Proposition 2.5.1. *[Chinese Remainder Theorem for simultaneous congruences] Let* $m, n \in \mathbb{N}$ *be coprime. Then for every pair of integers* a, b *the simultaneous congruences*

(2.5.1)
$$\begin{aligned} x &\equiv a \pmod m \\ x &\equiv b \pmod n \end{aligned}$$

have a solution which is unique modulo mn.

More generally, if $d = \gcd(m, n)$ *then the congruences (2.5.1) have a solution if and only if* $a \equiv b \pmod{d}$*, and the solution (when it exists) is unique modulo lcm*$(m, n) = mn/d$*.*

To find the solution in the coprime case, write $1 = mu + nv$. Then we have the solution $x = mub + nva$ since $nv \equiv 1 \pmod{m}, \equiv 0 \pmod{n}$ while $mu \equiv 0 \pmod{m}, \equiv 1 \pmod{n}$.

Example: Let $m = 13$, $n = 17$. Then $1 = \gcd(13, 17) = 52 - 51$ so the solution for general a, b is $x \equiv 52b - 51a \pmod{221}$.

The CRT says that there is a bijection between pairs $(a \mod m, b \mod n)$ and single residue classes $(c \mod mn)$ when m, n are coprime. This bijection is in fact a ring isomorphism:

Theorem 2.5.2. *[Chinese Remainder Theorem, algebraic form] Let* $m, n \in \mathbb{N}$ *be coprime. Then we have the isomorphism of rings*

$$\mathbb{Z}/mn\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

Restricting to units on both sides, we have the isomorphism of groups

$$U_{mn} \cong U_m \times U_n.$$

Both forms of the CRT extend to several moduli m_1, m_2, $\ldots$, m_k provided that they are *pairwise* coprime. The second part of the proposition has the following important corollary: φ is a *multiplicative function*.

Proposition 2.5.3. *Let* $m, n \in \mathbb{N}$ *be coprime. Then* $\varphi(mn) = \varphi(m)\varphi(n)$*.*

Corollary 2.5.4. *Let $m \in \mathbb{N}$ have prime factorization*

$$m = \prod_{i=1}^{k} p_i^{e_i}$$

where the p_i are distinct primes and $e_i \geq 1$. Then

$$\varphi(m) = \prod_{i=1}^{k} p_i^{e_i - 1}(p_i - 1) = m \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right).$$

Examples: (1). $\varphi(168) = \varphi(8)\varphi(3)\varphi(7)$ (splitting 168 into prime powers) $= (8-4)(3-1)(7-1) = 4 \cdot 2 \cdot 6 = 48$. Alternatively, $\varphi(168) = 168 \cdot \left(1 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{3}\right) \left(1 - \frac{1}{7}\right) = 168 \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{6}{7} = 48$.
(2). $\varphi(100) = \varphi(4)\varphi(25) = 2 \cdot 20 = 40$.

One more property of $\varphi(m)$ will be useful later.

Proposition 2.5.5. *Let $m \in \mathbb{N}$. Then $\sum_{d|m} \varphi(d) = m$.*

The sum here is over all positive divisors of m. For example, when $m = 12$ we have

$$\begin{aligned} 12 &= \varphi(1) + \varphi(2) + \varphi(3) + \varphi(4) + \varphi(6) + \varphi(12) \\ &= 1 + 1 + 2 + 2 + 2 + 4. \end{aligned}$$

Applications of CRT: The CRT says that congruences to two coprime moduli are, in a sense, independent. Solving a general congruence to a general modulus reduces to solving it modulo prime powers, and then using CRT to "glue" the separate solutions together.

For example: solve $x^2 \equiv 1 \pmod{91}$. Since $91 = 7 \cdot 13$ we first solve separately modulo 7 and modulo 13, giving $x \equiv \pm 1 \pmod 7$ and $x \equiv \pm 1 \pmod{13}$ by an earlier proposition since 7 and 13 are prime. This gives four possibilities modulo 91:

$$\begin{aligned}
(+1 \mod 7, \quad +1 \mod 13) &\leftrightarrow (+1 \mod 91) \\
(+1 \mod 7, \quad -1 \mod 13) &\leftrightarrow (-27 \mod 91) \\
(-1 \mod 7, \quad +1 \mod 13) &\leftrightarrow (+27 \mod 91) \\
(-1 \mod 7, \quad -1 \mod 13) &\leftrightarrow (-1 \mod 91)
\end{aligned}$$

So the solutions are $x \equiv \pm 1 \pmod{91}$ and $x \equiv \pm 27 \pmod{91}$. To solve the second and third we use the method given above: write $1 = 7u + 13v = 14 - 13$, then $(a, b) = (1, -1)$ maps to $mub + nva = 14b - 13a = 14(-1) - 13(1) \equiv -27 \pmod{91}$.

Systematic study of various types of congruence now follows the following pattern. First work modulo primes; this is easiest since $\mathbb{Z}/p\mathbb{Z}$ is a field. Then somehow go from primes to prime powers. The process here (called "Hensel lifting") is rather like taking successive decimal approximations to an ordinary equation, and we will come back to this at the end of the module, in Chapter 5 on p-adic numbers. Finally, use the CRT to "glue" together the information from the separate prime powers.

2.6. **The structure of U_m.** The most important result here is that for prime p, the multiplicative group U_p $(= \mathbb{F}_p^*)$ is cyclic.

Theorem 2.6.1. *Let p be a prime. Then the group $U_p = \mathbb{F}_p^*$ is cyclic.*

Definition 2.6.2. *An integer which generates $U_p = \mathbb{F}_p^*$ is called a* primitive root modulo p*. If U_m is cyclic, then a generator of U_m is called a* primitive root modulo m*.*

When g is a primitive root modulo m, the powers $1, g, g^2, \dots, g^{\varphi(m)-1}$ are incongruent modulo m, and every integer which is coprime to m is congruent to exactly one of these. The other primitive roots are the g^k for which $\gcd(k, \varphi(m)) = 1$. So we have the following:

Corollary 2.6.3. *Let p be a prime. Then p has a primitive root, and the number of incongruent primitive roots modulo p is $\varphi(p-1)$. More generally, for every $d|(p-1)$ there are $\varphi(d)$ integers (incongruent modulo p) with order d modulo p.*

If m has a primitive root then there are $\varphi(\varphi(m))$ incongruent primitive roots modulo m.

Example: Let $p = 13$. Since $\varphi(p-1) = \varphi(12) = 4$ there are 4 primitive roots modulo 13. One is 2, since the successive powers of 2 modulo 13 are $1, 2, 4, 8, 3, 6, -1, \dots$. The others are the powers 2^k where $\gcd(k, 12) = 1$: taking $k = 1, 5, 7, 11$ gives the primitive roots $2, 2^5 \equiv 6, 2^7 \equiv 11, 2^{11} \equiv 7 \pmod{13}$.

As an application of primitive roots, we may give a simple proof of a result proved earlier, that when $p \equiv 1 \pmod 4$ then the congruence $x^2 \equiv -1 \pmod p$ has a solution. For let g be a primitive root modulo p, and set $a = g^{(p-1)/4}$. Then $a^2 \equiv g^{(p-1)/2} \not\equiv 1 \pmod p$, but $a^4 = g^{p-1} \equiv 1 \pmod p$, from which it follows that $a^2 \equiv -1 \pmod p$.

Theorem 2.6.4. *Primitive roots modulo m exist if and only if $m = 1, 2, 4$, p^e or $2p^e$ where p is an odd prime and $e \geq 1$.*

Now if m is odd, with prime factorization $m = \prod_{i=1}^k p_i^{e_i}$, it follows that the group U_m is isomorphic to the product of cyclic groups of order $p_i^{e_i - 1}(p_i - 1)$ for $1 \leq i \leq k$.

We have not determined the structure of U_{2^e} for $e \geq 3$; it turns out that while not cyclic, it is almost so: for $e \geq 3$, U_{2^e} is isomorphic to the product of cyclic groups of order 2 (generated by -1) and order 2^{e-2} (generated by 5).

3. Quadratic Reciprocity

In this section we will study quadratic congruences to prime moduli. When p is an odd prime, then any quadratic congruence $ax^2 + bx + c \equiv 0 \pmod{p}$ (with $p \nmid a$) may be reduced by completing the square to the simpler congruence $y^2 \equiv d \pmod{p}$, where $d = b^2 - 4ac$ and $y = 2ax + b$. So solving quadratic congruences reduces to the problem of taking square roots.

3.1. Quadratic Residues and Nonresidues.

Definition 3.1.1. *Let p be an odd prime and a an integer not divisible by p. We say that a is a* quadratic residue *of p when $x^2 \equiv a \pmod{p}$ has at least one solution, and a* quadratic nonresidue *otherwise.*

Note that when a is a quadratic residue with $b^2 \equiv a \pmod{p}$ then the congruence $x^2 \equiv a \pmod{p}$ has exactly two solutions, namely $x \equiv \pm b$. For these are both solutions; they are incongruent modulo p since $b \equiv -b \implies 2b \equiv 0 \implies b \equiv 0 \implies a \equiv 0$. (Here we used that $p \neq 2$.) Lastly, there are no more solutions since $p|x^2 - a \implies p|x^2 - b^2 \implies p|(x-b)(x+b) \implies p|(x-b)$ or $p|(x+b)$.

We can find the quadratic residues modulo p by reducing b^2 modulo p for $1 \leq b \leq (p-1)/2$. The other squares will repeat these (in reverse order), since $(p-b)^2 \equiv b^2 \pmod{p}$. It follows that exactly half the nonzero residues are quadratic residues and the other half quadratic nonresidues.

Examples: $p = 11$: the quadratic residues modulo 11 are:

$$1^2, 2^2, 3^2, 4^2, 5^2 \equiv 1, 4, 9, 5, 3 \equiv 1, 4, -2, 5, 3$$

while the quadratic nonresidues are $2, 6, 7, 8, 10 \equiv 2, -5, -4, -3, -1$.

$p = 13$: the quadratic residues modulo 13 are:

$$1^2, 2^2, 3^2, 4^2, 5^2, 6^2 \equiv 1, 4, 9, 3, 12, 10 \equiv 1, 4, -4, 3, -1, -3$$

while the quadratic nonresidues are ± 2, ± 5, ± 6.

The reason for the patterns we see here will become apparent later.

Another way to see that exactly half the nonzero residues are quadratic residues is to use primitive roots. Let g be a primitive root modulo p. Then the nonzero residues are g^k for $0 \le k \le p-2$ and every integer not divisible by p is congruent to g^k for some k in this range. The quadratic residues are the g^k for even k: that is, the powers of g^2.

For example when $p = 13$ we may take $g = 2$, so $g^2 = 4$ with successive powers $1, 4, 3, 12, 9, 10 \pmod{13}$. These are the quadratic residues; to get the quadratic nonresidues multiply them by $g = 2$ to get the odd powers $2, 8, 6, 11, 5, 7 \pmod{13}$.

3.2. **Legendre Symbols and Euler's Criterion.**

Definition 3.2.1. *The* Legendre Symbol $\left(\dfrac{a}{p}\right)$ *is defined as follows:*

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & \text{if } p \nmid a \text{ and } x^2 \equiv a \pmod{p} \text{ has a solution} \\ -1 & \text{if } p \nmid a \text{ and } x^2 \equiv a \pmod{p} \text{ does not have a solution} \\ 0 & \text{if } p \mid a \end{cases}$$

In all cases, the number of (incongruent) solutions to $x^2 \equiv a \pmod{p}$ is $1 + \left(\dfrac{a}{p}\right)$.

Proposition 3.2.2. *Let p be an odd prime.*

(a) $a \equiv b \pmod{p} \implies \left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

(b) **Euler's Criterion:** $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$.

(c) $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} +1 & \text{if } p \equiv 1 \pmod{4} \\ -1 & \text{if } p \equiv 3 \pmod{4} \end{cases}$.

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

Corollary 3.2.3. *Let p be an odd prime.*

If $p \equiv 1 \pmod{4}$ then $\left(\frac{-a}{p}\right) = \left(\frac{a}{p}\right)$ for all a.

If $p \equiv 3 \pmod{4}$ then $\left(\frac{-a}{p}\right) = -\left(\frac{a}{p}\right)$ for all a.

If we start to ask questions such as "for which primes p is 2 a quadratic residue?" then we are led to one of the most famous results in elementary number theory. Experimental evidence for small primes easily convinces one that the answer is "primes congruent to $\pm 1 \pmod{8}$":

$$\left(\frac{2}{p}\right) = +1 \text{ for } p = 7, 17, 23, 31, 41, 47, 71, \ldots$$

$$\left(\frac{2}{p}\right) = -1 \text{ for } p = 3, 5, 11, 13, 19, 29, 37, 43, \ldots$$

More generally, the value of $\left(\frac{a}{p}\right)$ for fixed a and variable p only depends on the residue of p modulo $4a$. This is one form of Gauss's famous Law of Quadratic Reciprocity.

3.3. The Law of Quadratic Reciprocity.

Proposition 3.3.1. *[Gauss's Lemma] Let p be an odd prime and a an integer not divisible by p. Then $\left(\frac{a}{p}\right) = (-1)^s$, where s is the number of integers i with $0 < i < p/2$ for which the least residue of ai is negative.*

Example: Take $p = 13$ and $a = 11$; then we reduce $11, 22, 33, 44, 55, 66$ modulo 13 to $-2, -4, -6, 5, 3, 1$. As expected by the proof of the Proposition, these are, up to sign, the integers between 1 and 6. There are 3 minus signs, so $\left(\frac{11}{13}\right) = (-1)^3 = -1$.

If $p = 13$ and $a = 10$ then we reduce $10, 20, 30, 40, 50, 60$ to $-3, -6, 4, 1, -2, -5$ with four negative values, so $\left(\frac{10}{13}\right) = (-1)^4 = 1$. Indeed, $6^2 = 36 \equiv 10 \pmod{13}$.

Corollary 3.3.2. *Assume that $a > 0$, and set $a' = a$ if a is even, $a' = a - 1$ if a is odd. Then $\left(\frac{a}{p}\right) = (-1)^s$ where*

$$s = \sum_{k=1}^{a'} \left[(kp)/(2a)\right].$$

Example: Take $p = 13$ and $a = 11$, so $a' = 10$. Then $\left(\frac{11}{13}\right) = (-1)^s$ where $s = [13/22] + [26/22] + [39/22] + [52/22] + [65/22] + [78/22] + [91/22] + [104/22] + [117/22] + [130/22] \equiv 0 + (1+1) + (2+2) + 3 + (4+4) + (5+5) \equiv 1 \pmod 2$, so $\left(\frac{11}{13}\right) = -1$.

We can use Corollary 3.3.2 to Gauss's Lemma to evaluate $\left(\frac{2}{p}\right)$ for *all* odd primes p.

Proposition 3.3.3. *Let p be an odd prime. Then*

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} +1 & \text{if } p \equiv \pm 1 \pmod 8; \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

More generally, we can deduce that in general the value of $\left(\frac{a}{p}\right)$ only depends on $p \pmod{4a}$, our first form of *quadratic reciprocity*: although the definition of $\left(\frac{a}{p}\right)$ is in terms of $a \pmod p$, it is far from obvious that it depends on $p \pmod{4a}$!

Proposition 3.3.4. *Let p and q be odd primes and a a* positive *integer not divisible by either p or q. Then*

$$p \equiv \pm q \pmod{4a} \implies \left(\frac{a}{p}\right) = \left(\frac{a}{q}\right).$$

(For $a < 0$ a slightly modified result holds: exercise.)

The *Law of Quadratic Reciprocity* uses this result in the case that a is also prime to get a very symmetric statement.

Theorem 3.3.5. *[Quadratic Reciprocity] Let p and q be distinct odd primes. Then*

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

So $\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)$ *if* $p \equiv 1$ *or* $q \equiv 1 \pmod 4$, *while* $\left(\frac{q}{p}\right) = -\left(\frac{p}{q}\right)$ *if* $p \equiv q \equiv 3 \pmod 4$.

Since the Legendre symbol $\left(\frac{a}{p}\right)$ is completely multiplicative in a for fixed p, to evaluate $\left(\frac{a}{p}\right)$ for all a we only need to know the values of $\left(\frac{-1}{p}\right)$, $\left(\frac{2}{p}\right)$ and $\left(\frac{q}{p}\right)$, for odd primes q different from p. The Law of Quadratic Reciprocity tells us how to evaluate each of these! Special cases of the reciprocity law were conjectured by Euler on the basis of substantial calculations and knowledge, but Gauss first proved it, and in fact gave several proofs.

Summary of Quadratic Reciprocity: If p and q are distinct odd primes then:

- $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} +1 & \text{if } p \equiv 1 \pmod 4; \\ -1 & \text{if } p \equiv 3 \pmod 4; \end{cases}$
- $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} +1 & \text{if } p \equiv \pm 1 \pmod 8; \\ -1 & \text{if } p \equiv \pm 3 \pmod 8; \end{cases}$

- $\left(\frac{q}{p}\right) = \begin{cases} +\left(\frac{p}{q}\right) & \text{if } \textit{either } p \equiv 1 \pmod 4 \ \textit{ or } q \equiv 1 \pmod 4; \\ -\left(\frac{p}{q}\right) & \text{if } \textit{both } p \equiv 3 \pmod 4 \ \textit{ and } q \equiv 3 \pmod 4. \end{cases}$

Using QR we may easily answer questions of the form: Given a, for which p is $\left(\frac{a}{p}\right) = 1$? For example:

$$\left(\frac{-2}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{2}{p}\right) = \begin{cases} +1 & \text{if } p \equiv 1, 3 \pmod 8; \\ -1 & \text{if } p \equiv -1, -3 \pmod 8. \end{cases}$$

$$\left(\frac{-3}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right) = \begin{cases} +1 & \text{if } p \equiv 1 \pmod 3; \\ -1 & \text{if } p \equiv -1 \pmod 3. \end{cases}$$

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

(Notice how $\left(\frac{a}{p}\right)$ sometimes depends only on p modulo a rather than modulo $4a$.)

Using Proposition 3.3.4 gives an alternative method of evaluating $\left(\frac{a}{p}\right)$ for fixed $a > 0$. Take $a = 3$, so we know that $\left(\frac{3}{p}\right)$ only depends on $\pm p \pmod{12}$; when $p = 13$ we have $\left(\frac{3}{13}\right) = +1$

and when $p = 5$ we have $\left(\frac{3}{5}\right) = -1$; so $\left(\frac{3}{p}\right) = +1$ for all $p \equiv \pm 1 \pmod{12}$ and $\left(\frac{3}{p}\right) = -1$ for all $p \equiv \pm 5 \pmod{12}$.

When $a < 0$ it is also true that $p \equiv q \pmod{4a} \implies \left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$, but now $p \equiv -q \pmod{4a} \implies \left(\frac{a}{p}\right) = -\left(\frac{a}{q}\right)$. (Apply Prop. 3.3.4 to $-a$ to see this.) Hence we can evaluate $\left(\frac{a}{p}\right)$ for $a < 0$.

For example, take $a = -5$. Then $\left(\frac{-5}{p}\right)$ depends on p modulo 20, giving $\varphi(20) = 8$ cases. Take the primes $p = 61, 3, 7, 29$ which are congruent respectively to $1, 3, 7, 9 \pmod{20}$; computing the four Legendre symbols $\left(\frac{-5}{p}\right)$, we find that they are all $+1$. Hence

$$\left(\frac{-5}{p}\right) = \begin{cases} +1 & \text{if } p \equiv 1, 3, 7, 9 \pmod{20}; \\ -1 & \text{if } p \equiv 11, 13, 17, 19 \pmod{20}. \end{cases}$$

where the second line follows from the first by the "anti-symmetry" since $-5 < 0$.

4. Diophantine Equations

A *Diophantine Equation* is simply an equation in one or more variables for which *integer* (or sometimes rational) solutions are sought. For example:

- $x^2 + y^2 = z^2$ has solutions $(x, y, z) = (3, 4, 5), (5, 12, 13), \dots;$
- $x^3 + y^3 = z^3$ has no solutions with x, y, z positive integers;
- $x^2 - 61y^2 = 1$ has infinitely many solutions with $x, y > 0$; the smallest has $x = 1766319049$ and $y = 226153980$.

We will use the techniques we have developed in earlier chapters, as well as one new one, to solve a number of Diophantine equations all of which have had some historical interest. Their solution has led to the development of much of modern algebra and number theory. The new technique we will use is called the *Geometry of Numbers*.

4.1. Geometry of Numbers and Minkowski's Theorem.

We will use the geometry of $\mathbb{R}^n$ and of certain subsets of it:

Definition 4.1.1. *A* lattice *in $\mathbb{Z}^n$ is a subgroup $L \subseteq \mathbb{Z}^n$ of finite index.*

The lattices we will use are all defined using congruence conditions on the coordinates of vectors in $\mathbb{Z}^n$, and the index of the lattice will be determined from the moduli of these congruences (example to follow soon). There are more general subsets of $\mathbb{R}^n$ called lattices, but we will not need them.

Our general strategy will be to set up a lattice so that the coordinates give a "modular approximation" to the equation being solved; then to get an exact solution we require a second condition, that the vector of coefficients is "small" in some sense. Minkowski's Theorem will show that (under certain conditions) there are short lattice vectors, and we win. Its statement requires the following definitions.

Definition 4.1.2. *A subset $S \subseteq \mathbb{R}^n$ is* symmetric *if $x \in S \iff -x \in S$, and* convex *if $x, y \in S \implies tx + (1-t)y \in S$ for all t with $0 \leq t \leq 1$.*

Here is the result from the geometry of numbers we will use to deduce the existence of solutions to several Diophantine Equations:

Theorem 4.1.3. *[Minkowski] Let $L \leq \mathbb{Z}^n$ be a lattice of index m, and let $S \subseteq \mathbb{R}^n$ be a bounded convex symmetric domain. If S has volume $v(S) > 2^n m$, then S contains a nonzero element of L.*

The same conclusion holds when $v(S) = 2^n m$, provided that S is compact.

4.2. **Sums of squares.** In this section we will give an answer to the questions "which positive integers can be expressed as a sum of 2 squares (S2S), or a sum of 3 squares (S3S), or a sum of 4 squares (S4S)"? In the 3-squares case we will only give a partial proof, since the full proof uses concepts which we will not cover. The reason for the S3S case being harder is that the set of S3S numbers is not closed under multiplication, while for S2S and S4S it is, which then essentially reduces the question to the case of primes.

4.2.1. *Sums of two squares.* To ask whether an integer n is a sum of two squares, $n = a^2 + b^2$, is the same as to ask whether it is the norm of a Gaussian Integer: $n = a^2 + b^2 = N(\alpha)$ where $\alpha = a + bi \in \mathbb{Z}[i]$. Using Theorem 1.5.14 on Gaussian primes, such an integer must be a product of norms of Gaussian primes which are: 2, p for any prime $p \equiv 1 \pmod 4$, and q^2 for any prime $q \equiv 3 \pmod 4$. This proves the following:

Theorem 4.2.1. *The positive integer n may be expressed as a sum of two squares, $n = x^2 + y^2$, if and only if $\mathrm{ord}_q(n)$ is even for all primes $q \equiv 3 \pmod 4$, or equivalently if and only if $n = ab^2$ where a has no prime factors congruent to $3 \pmod 4$.*

Remarks: One can similarly characterize positive integers of the form $n = x^2 + 2y^2$ as those such that $\mathrm{ord}_q(n)$ is even for all primes $q \equiv 5, 7 \pmod 8$. Either a direct proof or one based on unique factorization in the Euclidean Domain $\mathbb{Z}[\sqrt{-2}]$ is possible. A similar result holds for $n = x^2 + 3y^2$ (though is slightly harder to prove since $\mathbb{Z}[\sqrt{-3}]$ is not Euclidean). But the pattern does not continue, and for general m it is a very hard problem to determine exactly which integers n, or even which primes p, have the form $x^2 + my^2$. The study of this question leads on to algebraic number theory, and in particular to the study of the arithmetic properties of quadratic number fields.

Recall from Chapter 1 that the key to determining the Gaussian primes was a fact which we only proved later (Theorem 2.4.2): that if p is a prime such that $p \equiv 1 \pmod 4$ then p is a sum of two squares. We proved this in Chapter 2 by using facts about Gaussian Integers, together with the fact that for such primes the congruence $x^2 \equiv -1 \pmod p$ has a solution. Now we give a different proof that $p \equiv 1 \pmod 4 \implies p = a^2 + b^2$, as a first application of the Geometry of Numbers.

Theorem 4.2.2. *[=Theorem 2.4.2 again] Let p be a prime such that $p \equiv 1 \pmod 4$. Then there exist integers a and b such that $p = a^2 + b^2$.*

Before applying Minkowski again to prove the four-square theorem below, we will briefly (and incompletely) look at sums of three squares.

4.2.2. *Sums of three squares.*

Proposition 4.2.3. *Let n be a positive integer with $n \equiv 7 \pmod 8$. Then n is not a sum of three squares, and nor is any integer of the form $4^k n$ with $n \equiv 7 \pmod 8$.*

The converse of this result is true: every positive integer not of the form $4^k n$ with $n \equiv 7 \pmod 8$ can be written as a sum of three squares. But this is harder to prove and we omit it. Instead we turn to sums of four squares.

4.2.3. *Sums of four squares.*

Theorem 4.2.4. *[Lagrange] Every positive integer may be expressed as a sum of four squares.*

Note that 0 is allowed as one of the squares. The theorem will follow from the following Lemma 4.2.5, which reduces the problem to expressing all primes as S4S, and Proposition 4.2.6 which shows that all primes are S4S.

Lemma 4.2.5. *If $m = a_1^2 + a_2^2 + a_3^2 + a_4^2$ and $n = b_1^2 + b_2^2 + b_3^2 + b_4^2$ then $mn = c_1^2 + c_2^2 + c_3^2 + c_4^2$ where*

$$\begin{aligned}
c_1 &= a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4 \\
c_2 &= a_1b_2 - a_2b_1 + a_3b_4 - a_4b_3 \\
c_3 &= a_1b_3 - a_3b_1 - a_2b_4 + a_4b_2 \\
c_4 &= a_1b_4 - a_4b_1 + a_2b_3 - a_3b_2.
\end{aligned}$$

Proposition 4.2.6. *Every prime number may be expressed as a sum of four squares.*

4.3. **Legendre's Equation.** Here is an example of an equation with **no** nontrivial solutions.

Example: The equation $x^2 + y^2 = 3z^2$ has no integer solutions except $x = y = z = 0$.

For suppose that (x, y, z) is a nonzero solution. Then we may assume that $\gcd(x, y) = 1$ since if both x and y were divisible by some prime p, then $p^2|3z^2$ and so $p|z$, so we could divide through by p^2 to get the smaller nontrivial solution $(x/p, y/p, z/p)$. Next, neither x nor y is divisible by 3 (since if either is then so would the other be). This implies $x \equiv \pm 1 \pmod 3$ and $y \equiv \pm 1 \pmod 3$, so $x^2 + y^2 \equiv 1 + 1 = 2 \not\equiv 0 \pmod 3$, contradicting $x^2 + y^2 = 3z^2$.

We have used two properties of the number 3 here: that it is square-free (so $p^2|3z^2 \implies p|z$) and that $x^2 + y^2 \equiv 0 \pmod 3 \implies x \equiv y \equiv 0 \pmod 3$. So the same argument works for the equations $x^2 + y^2 = qz^2$ where q is any prime congruent to $3 \pmod 4$.

The general equation

$$ax^2 + by^2 = cz^2 \tag{4.3.1}$$

with $a, b, c \in \mathbb{N}$ has been studied since the 19th century, and is known as *Legendre's Equation*. There is a simple criterion for the existence of nontrivial solutions in terms of congruences modulo a, b and c. By a *solution* to (4.3.1) we will always mean a solution other than the trivial one $(x, y, z) = (0, 0, 0)$. By homogeneity, (x, y, z) satisfies (4.3.1) if and only if (rx, ry, rz) also does for any $r \neq 0$; a solution will be called *primitive* if $\gcd(x, y, z) = 1$.

First we reduce to the case where a, b, c are pairwise coprime and square-free:

- If $d = \gcd(a, b) > 1$ then (x, y, z) satisfies (4.3.1) if and only if (dx, dy, z) satisfies the similar equation with coefficients $(a/d, b/d, cd)$. Similarly if $\gcd(a, c) > 1$ or $\gcd(b, c) > 1$. Note that the product abc is reduced (by a factor d) in each case, so after a finite number of such steps we may assume that a, b, c are pairwise coprime.

- If $d^2|a$ then (x, y, z) satisfies (4.3.1) if and only if (dx, y, z) satisfies the similar equation with coefficients $(a/d^2, b, c)$. Similarly with square factors of b or c, so we can assume that each of a, b, c is square-free.

Theorem 4.3.1. *Let $a, b, c \in \mathbb{N}$ be pairwise coprime and square-free. Then a non-trivial solution to (4.3.1) exists if and only if each of the quadratic congruences*

$$x^2 \equiv bc \pmod{a}, \qquad x^2 \equiv ac \pmod{b}, \qquad x^2 \equiv -ab \pmod{c}$$

has a solution.

Our proof just fails to show that there always is a solution satisfying the inequalities $|x| \leq \sqrt{bc}$, $|y| \leq \sqrt{ac}$, $|z| \leq \sqrt{ab}$, because of the adjustment needed at the end; however there is always such a "small" solution (proof omitted).

To make the proof constructive, we would need to have a method for finding short vectors in lattices. Such methods do exist (the most famous is the LLL method named after Lenstra, Lenstra and Lovasz) and have a huge number of applications in computational number theory and cryptography. One reason that lattice-based methods are becoming popular in cryptography is that they are "quantum-resistant", meaning that no-one (yet!) knows how to solve problems such as the SVP (Shortest Vector Problem) using a quantum computer, unlike the case for factorization-based methods such as RSA.

4.4. **Pythagorean Triples.** A classical problem is to find all right-angled triangles all of whose sides have integral length. Letting the sides be x, y and z this amounts (by Pythagoras's Theorem) to finding positive integer solutions to the Diophantine equation

$$x^2 + y^2 = z^2. \tag{4.4.1}$$

A solution (x, y, z) is called a *Pythagorean Triple*. For example, $(3, 4, 5)$ is a Pythagorean Triple.

Clearly if (x, y, z) is a Pythagorean Triple then so is (kx, ky, kz) for all $k \geq 1$, and to avoid this trivial repetition of solutions we will restrict to *Primitive Pythagorean Triples* which have the additional property that $\gcd(x, y, z) = 1$. From (4.4.1) it then follows that x, y, z are pairwise coprime, since a prime divisor of any two would have to divide the third.

Finally, in any primitive Pythagorean Triple, exactly one of x and y is even, the other odd; for they are not both even by primitivity, and cannot both be odd for then $x^2 + y^2 \equiv 2 \pmod 4$, so $x^2 + y^2$ could not be a square. By symmetry we only consider triples with x and z odd, y even.

The following result shows how to parametrize all primitive Pythagorean Triples.

Theorem 4.4.1. *Let u and v be positive coprime integers with $u \not\equiv v \pmod 2$ and $u > v$. Set*

$$x = u^2 - v^2; \qquad y = 2uv; \qquad z = u^2 + v^2.$$

Then (x, y, z) is a primitive Pythagorean Triple. Conversely, all primitive Pythagorean Triples are obtained in this way for suitable u and v.

We will see an application of our parametrization of Pythagorean triples to the Fermat equation $x^4 + y^4 = z^4$ in the next section. This case of Fermat's Last Theorem says that there are no Pythagorean Triples with all three integers perfect squares.

An alternative approach to the previous Theorem is to use the Gaussian Integers $\mathbb{Z}[i]$. Suppose $x^2 + y^2 = z^2$ with $\gcd(x, y) = 1$ and z odd. Then $z^2 = (x + yi)(x - yi)$, and the factors on the right are coprime: for if $\alpha | x+yi$ and $\alpha | x-yi$ for some $\alpha \in \mathbb{Z}[i]$, then $\alpha | 2x$ and $\alpha | 2yi$, from which $\alpha | 2$ since $\gcd(x,y)=1$ and i is a unit. But $\gcd(z,2)=1$ so α is a unit.

Now each of $x \pm yi$ must be a square or a unit times a square, since they are coprime and their product is a square **and $\mathbb{Z}[i]$ is a UFD**. If $x+yi = \pm(u+vi)^2$ then $x = \pm(u^2-v^2)$ and $y = \pm 2uv$; if $x+yi = \pm i(u+vi)^2$ then $x = \mp 2uv$ and $y = \pm(u^2-v^2)$. The proof that $\gcd(u,v)=1$ and $u \not\equiv v \pmod 2$ is as before, or follows from the fact that $u+vi$ and $u-vi$ are coprime in $\mathbb{Z}[i]$.

Other similar equations may be solved by the same method. For example, all primitive solutions to $x^2+2y^2=z^2$ are obtained from $(x,y,z) = (\pm(u^2-2v^2), \pm 2uv, \pm(u^2+2v^2))$. This can be proved using the UFD $\mathbb{Z}[\sqrt{-2}]$ or by elementary means.

4.5. **Fermat's Last Theorem.** After our success in finding all solutions to the equation $x^2+y^2=z^2$, it is natural to turn to analogous equation for higher powers. So we ask for solutions in positive integers to the equation

$$x^n + y^n = z^n \qquad \text{with } n \geq 3. \tag{4.5.1}$$

Fermat claimed, in the famous marginal note to his edition of the works of Diophantus, that there are no solutions to (4.5.1). The result is known as *Fermat's Last Theorem*: it is the last of Fermat's unproved claims to be proved (or disproved). Since 1994 it has become possible to state the result as a Theorem:

Theorem 4.5.1. *[Fermat's Last Theorem; Wiles and Taylor–Wiles, 1994] Let $n \geq 3$. Then there are no solutions in positive integers to the equation $x^n+y^n=z^n$.*

The only case which we know that Fermat proved is $n = 4$, which we will prove below. Euler proved the case $n = 3$, using arithmetic in the ring $\mathbb{Z}[\sqrt{-3}]$, though there is some doubt as to the validity of Euler's argument at a crucial step where he tacitly assumed that this ring had unique factorization (which it does not). Subsequent work by Dirichlet, Legendre, Kummer and many others settled many more exponents, at the same time creating most of modern algebraic number theory and algebra. By 1987, the Theorem was known to be true for all $n \leq 150000$. In 1986, an unexpected connection was found, by Frey, between the Fermat equation and another class of Diophantine equation called *Elliptic curves*. A solution to Fermat's equation would lead to the existence of an elliptic curve with properties so strange that they would contradict widely-believed, but then unproved, conjectures about elliptic curves. This connection was proved by Ribet. Finally, Andrew Wiles, with the help of Richard Taylor, proved the elliptic curve conjecture, firmly establishing the truth of Fermat's Last theorem.

We will prove the case $n = 4$ of the theorem.

Theorem 4.5.2. *[Fermat's Last Theorem for exponent 4] The equation $x^4 + y^4 = z^4$ has no solutions in positive integers.*

We will prove a stronger statement: $x^4 + y^4$ cannot be a square, let alone a 4th power:

Theorem 4.5.3. *The equation $x^4 + y^4 = z^2$ has no solutions in positive integers.*

Corollary 4.5.4. *Let $n \in \mathbb{N}$ be a multiple of 4. Then there are no solutions in positive integers to the equation $x^n + y^n = z^n$.*

Now to prove Fermat's Last Theorem in general it suffices to show that the equation $x^p + y^p = z^p$ has no positive integer solutions for each *odd prime* p, since every $n \geq 3$ is divisible either by 4 or by an odd prime, and impossibility for a divisor of n implies impossibility for n itself.

4.6. **Proof of Minkowski's Theorem.** There are several ways to prove Minkowski's Theorem 4.1.3, all of which are based on a continuous analogue of the pigeon-hole principle. We'll use a preliminary result called Blichfeld's Theorem:

Theorem 4.6.1. *[Blichfeld's theorem] Let S be a bounded subset of $\mathbb{R}^n$ whose volume $v(S)$ exists and satisfies $v(S) > m$ for some integer $m \geq 1$. Then there exist $m + 1$ distinct points $\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_m \in S$ such that $\underline{x}_i - \underline{x}_j \in \mathbb{Z}^n$ for all i, j.*

5. p-adic Numbers

5.1. Motivating examples. We all know that $\sqrt{2}$ is irrational, so that 2 is not a square in the rational field $\mathbb{Q}$, but that we can enlarge $\mathbb{Q}$ to the real field $\mathbb{R}$ where 2 is a square. In $\mathbb{R}$, we may represent irrational numbers by (non-terminating, non-recurring) decimal expansions:

$$\sqrt{2} = 1.414213562373\cdots = 1 + 4\cdot 10^{-1} + 1\cdot 10^{-2} + 4\cdot 10^{-3} + 2\cdot 10^{-4} + \dots$$

In general, real numbers are expressible as

$$x = \pm \sum_{k=-\infty}^{n} a_k 10^k,$$

where the digits $a_k \in \{0,1,2,3,4,5,6,7,8,9\}$; there are only finitely many terms with $k>0$, but may be infinitely many with $k<0$; the series always converges in $\mathbb{R}$; and the sequence of digits (a_k) is usually uniquely determined by x. (The exceptions are numbers x with finite decimal expansions, where we can replace the tail $\dots a000\dots$ with $\dots(a-1)999\dots$.)

Another way of thinking about the decimal expansion of the irrational number $\sqrt{2}$ is to say that $\sqrt{2}$ is the limit of a sequence (x_k) of rational numbers: $x_0 = 1$, $x_1 = 14/10$, $x_2 = 141/100$, $\dots$. This is a Cauchy sequence of rational numbers, and has no limit in $\mathbb{Q}$, but does have a limit $\sqrt{2} = \lim_{k\to\infty} x_k$ in the larger complete field $\mathbb{R}$. The rational numbers x_k are rational approximations to $\sqrt{2}$, each being a better approximation than the previous one:

$$|\sqrt{2} - x_k| \le 10^{-k}.$$

As a first example of a p-adic number for $p = 7$, we consider the quadratic congruences

$$x^2 \equiv 2 \pmod{7^k}$$

for $k = 1, 2, 3 \ldots$. When $k = 1$ there are two solutions: $x = x_1 \equiv \pm 3 \pmod 7$. Any solution x_2 to the congruence modulo 7^2 must also be a solution modulo 7, hence of the form $x_2 = x_1 + 7y = \pm 3 + 7y$; choosing $x_1 = 3$ gives $x_2 = 3 + 7y$, which must satisfy

$$0 \equiv x_2^2 - 2 \equiv (3 + 7y)^2 - 2 \equiv 7(1 + 6y) \pmod{7^2};$$

equivalently, $1 + 6y \equiv 0 \pmod 7$ with unique solution $y \equiv 1 \pmod 7$; so $x_2 = 3 + 1 \cdot 7 = 10$.

Continuing in a similar way, setting $x_3 = x_2 + 7^2 y$ and substituting, we find that $x_3^2 \equiv 2 \pmod{7^3} \iff y \equiv 2 \pmod 7$, so $x_3 \equiv x_2 + 2 \cdot 7^2 \equiv 108 \pmod{7^3}$. The process may be continued indefinitely. At each stage there is a unique solution, so (after fixing the initial choice of $x_1 = 3$) we find, uniquely,

$$\begin{aligned} x_1 &= 3 = 3, \\ x_2 &= 10 = 3 + 1 \cdot 7, \\ x_3 &= 108 = 3 + 1 \cdot 7 + 2 \cdot 7^2, \\ x_4 &= 2166 = 3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3, \ldots \end{aligned}$$

The general formula is $x_{k+1} \equiv x_k^2 + x_k - 2 \pmod{7^{k+1}}$.

What happens "in the limit"? Does it even make sense to talk about the limit of the sequence x_k? Certainly there can be no *single* integer x satisfying $x^2 \equiv 2 \pmod{7^n}$ simultaneously for all $n \geq 1$, for then $x^2 - 2$ would be divisible by arbitrarily large powers of 7 which is only possible when $x^2 - 2 = 0$. Also, the infinite series $3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + \ldots$ does not converge in the normal sense, since the successive terms do not tend to 0.

We will define a new kind of number called a p-adic number, for each prime p. The p-adic integers will form a ring $\mathbb{Z}_p$, which contains $\mathbb{Z}$; there is one such ring for each prime p. In the ring $\mathbb{Z}_7$ of 7-adic integers, our sequence $3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + \dots$ will converge to a 7-adic limit, so that the equation $x^2 = 2$ has a solution in $\mathbb{Z}_7$. The solution can be expressed as an infinite 7-adic expansion:

$$\begin{aligned} x &= 3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + 7^4 + 2 \cdot 7^5 + 7^6 + 2 \cdot 7^7 + 4 \cdot 7^8 + 6 \cdot 7^9 + \dots \\ &= \sum_{k=0}^{\infty} a_k 7^k, \end{aligned}$$

where the "digits" a_k are all in the set $\{0, 1, 2, 3, 4, 5, 6\}$ and are uniquely determined after fixing $x \equiv 3 \pmod 7$: $a_0 = 3$, $a_1 = 1$, $a_2 = 2$, $a_3 = 6$, $\dots$.

The ring $\mathbb{Z}_p$ has a field of fractions $\mathbb{Q}_p$, which contains the rational field $\mathbb{Q}$. In fact, $\mathbb{Q}_p$ may be constructed directly from $\mathbb{Q}$ by a process similar to the construction of the real numbers as the set of limits of Cauchy sequences of rationals. $\mathbb{R}$ is the completion of $\mathbb{Q}$, complete in the usual analytic sense that Cauchy sequences converge in $\mathbb{R}$. Just as one can define the real numbers as (equivalence classes of) Cauchy sequences of rational numbers, we will start by defining p-adic integers as equivalence classes of suitable sequences of ordinary integers.

5.2. **Definition of $\mathbb{Z}_p$.** Fix, once and for all, a prime number p.

Definition 5.2.1. *A p-adic integer α is defined by a sequence of integers x_k for $k \geq 1$*

$$\alpha = \{x_k\}_{k=1}^{\infty} = \{x_1, x_2, x_3, \dots\},$$

satisfying the conditions

(5.2.1) $$x_{k+1} \equiv x_k \pmod{p^k} \qquad \textit{for all } k \geq 1,$$

with two sequences $\{x_k\}$ *and* $\{y_k\}$ *determining the same* p*-adic integer* α *if and only if*

$$x_k \equiv y_k \pmod{p^k} \qquad \textit{for all } k \geq 1.$$

The set of p*-adic integers is denoted* $\mathbb{Z}_p$*.*

An integer sequence satisfying (5.2.1) will be called *coherent*. Thus, each p-adic integer is actually an equivalence class of coherent sequences of ordinary integers, any one of which may be used to represent it. The representation of a p-adic integer $x = \{x_k\}$ will be called *reduced* if $0 \leq x_k < p^k$ for all $k \geq 1$. Every p-adic integer has a unique reduced representation.

The ordinary integers $\mathbb{Z}$ embed into $\mathbb{Z}_p$ as constant sequences, via $x \mapsto \{x, x, x, \dots\}$; this map is injective since if $x, y \in \mathbb{Z}$ satisfy $x \equiv y \pmod{p^k}$ for all $k \geq 1$, then $x = y$. So we can view $\mathbb{Z}$ as a subset of $\mathbb{Z}_p$. We may call elements of $\mathbb{Z}$ *rational integers* to distinguish them from p-adic integers.

Examples: Take $p = 3$. Here are three elements of $\mathbb{Z}_3$:

$$\begin{aligned} \alpha &= 40 = \{40, 40, 40, 40, 40, \dots\} = \{1, 4, 13, 40, 40, \dots\}; \\ \beta &= -1 = \{-1, -1, -1, -1, -1, \dots\} = \{2, 8, 26, 80, 242, \dots\}; \\ \gamma &= ? = \{1, 7, 16, 70, 151, \dots\}. \end{aligned}$$

the last representation is reduced in each case. Later we will see that γ is actually a representation of the rational number $-7/8$! In the reduced representation of -1, notice that

$$\begin{aligned} 2 &= 3 - 1 = 2, \\ 8 &= 3^2 - 1 = 2 + 2 \cdot 3, \\ 26 &= 3^3 - 1 = 2 + 2 \cdot 3 + 2 \cdot 3^2, \\ 80 &= 3^4 - 1 = 2 + 2 \cdot 3 + 2 \cdot 3^2 + 2 \cdot 3^3, \\ 242 &= 3^5 - 1 = 2 + 2 \cdot 3 + 2 \cdot 3^2 + 2 \cdot 3^3 + 2 \cdot 3^4, \end{aligned}$$

suggesting that the limiting value of the sequence x_k is $2(1 + 3 + 3^2 + 3^3 + \dots)$. This geometric series does not converge in the usual sense; but if it did converge, the usual formula would give as its sum the correct value $2/(1-3) = -1$. We will see later that this is a perfectly valid computation within the field $\mathbb{Q}_3$ of 3-adic numbers.

It follows from the coherence condition (5.2.1) that $\alpha = \{x_1, x_2, x_3, \dots\} = \{x_2, x_3, x_4, \dots\}$! In other words, we can shift the sequence any number of steps, or even delete any finite number of terms without affecting the value. At first sight this seems strange, but if you think of the value of α as being the *limit* of the sequence (x_k) (which we will later see to be the case), then it is natural.

We will see this index-shifting in action in proving some facts about p-adic numbers soon.

As suggested by the second example above, we now consider an alternative representation of a p-adic integer α with reduced representation $\{x_k\}$. Writing x_k to base p, we have

(5.2.2) $$x_k = a_0 + a_1 \cdot p + a_2 \cdot p^2 + \dots + a_{k-1} \cdot p^{k-1}$$

with each "digit" $a_i \in \{0, 1, 2, \dots, p-1\}$. The coherency condition (5.2.1) implies that $x_1 = a_0$, $x_2 = a_0 + a_1 p$, $x_3 = a_0 + a_1 p + a_2 p^2$, and so on, with the *same* digits a_i. So each $\alpha \in \mathbb{Z}_p$ determines a unique infinite sequence of *p-adic digits* $(a_i)_{i=0}^{\infty}$ with $0 \le a_i \le p-1$, and conversely every such digit sequence determines a unique p-adic integer $\alpha = \{x_k\}$ via (5.2.2). In the examples, the 3-adic digits of $\alpha = 40 = 1 + 3 + 3^2 + 3^3$ are $1, 1, 1, 1, 0, 0, \dots$ (effectively a finite sequence), those of $\beta = -1$ form the infinite recurring sequence $2, 2, 2, 2, 2, \dots$ and those of $\gamma = 1 + 2 \cdot 3 + 3^2 + 2 \cdot 3^3 + 3^4 + \dots$ are $1, 2, 1, 2, 1, \dots$.

We will write $\alpha = \{x_k\} = \sum_{i=0}^{\infty} a_i p^i$ when the p-adic digits of α are a_i, so that $x_k = \sum_{i=0}^{k-1} a_i p^i$ for $k \ge 1$. For now, this infinite series should be regarded as just a formal expression or shorthand.

5.3. **The ring $\mathbb{Z}_p$.** To add and multiply p-adic integers, just add and multiply the representative sequences termwise:

$$\{x_k\} + \{y_k\} = \{x_k + y_k\};$$
$$\{x_k\} \cdot \{y_k\} = \{x_k y_k\}.$$

One must check that the sequences on the right are coherent (in the sense of (5.2.1)), and that replacing $\{x_k\}$ or $\{y_k\}$ by an equivalent sequence does not change the equivalence classes of the sequences on the right: these are straightforward exercises, as are the verifications that all the ring axioms hold. For example, the negative of $\alpha = \{x_k\}$ is just $-\alpha = \{-x_k\}$. Expressing these operations in terms of the expansions $\alpha = \sum a_i p^i$ is not so easy: we will see examples later.

This gives $\mathbb{Z}_p$ the structure of a *commutative ring*, with $\mathbb{Z}$ as a subring. The factorization theory of p-adic integers turns out to be rather simple. There are no zero-divisors:

Proposition 5.3.1. *$\mathbb{Z}_p$ is an integral domain.*

Next we determine the units $U(\mathbb{Z}_p)$:

Proposition 5.3.2. *Let* $\alpha = \{x_k\} = \sum a_i p^i \in \mathbb{Z}_p$. *The following are equivalent:*

(i) $\alpha \in U(\mathbb{Z}_p)$;
(ii) $p \nmid x_1$;
(iii) $p \nmid x_k$ *for all* $k \geq 1$;
(iv) $a_0 \neq 0$;

Examples: If $a \in \mathbb{Z}$ with $p \nmid a$, then a is a p-adic unit. Its inverse is given by the coherent sequence $\{x_k\}$ where x_k satisfies $ax_k \equiv 1 \pmod{p^k}$ for $k \geq 1$.

For example, 3 is a 5-adic unit, so $1/3 \in \mathbb{Z}_5$. To find the terms x_k in its defining sequence for $k \leq 4$, solve $3x_4 \equiv 1 \pmod{5^4}$ to get $x_4 = 417$. Reducing this modulo lower powers of 5 then gives the start of the sequence in reduced form: $1/3 = \{2, 17, 42, 417, \dots\}$. And since $417 = 2 + 3 \cdot 5 + 5^2 + 3 \cdot 5^3$, the 5-adic digits of $1/3$ start $2, 3, 1, 3, \dots$. In fact the digit sequence recurs: $2, 3, 1, 3, 1, 3, 1, 3, 1, 3 \dots$. We can verify this by summing the series:

$$1 + (1 + 3 \cdot 5)(1 + 5^2 + 5^4 + \dots) = 1 + 16/(1 - 25) = (24 - 16)/24 = 1/3.$$

As another example, expanding $-7/8$ in $\mathbb{Z}_3$ gives the example denoted γ above (exercise).

It is easy to tell whether a p-adic integer is divisible by p, or by a power of p:

Proposition 5.3.3. *For* $\alpha = \{x_k\} \in \mathbb{Z}_p$:

(i) $p | \alpha \iff \alpha \notin U(\mathbb{Z}_p) \iff x_1 \equiv 0 \pmod{p} \iff x_k \equiv 0 \pmod{p}\ (\forall k \geq 1)$;
(ii) *for* $n \geq 1$, $p^n | \alpha \iff x_n \equiv 0 \pmod{p^n} \iff x_k \equiv 0 \pmod{p^n}\ (\forall k \geq n)$.

Now we know that every p-adic integer is either a unit or a multiple of p, but never both. From this we can show that $\mathbb{Z}_p$ is a UFD, with p the only prime:

Theorem 5.3.4. *$\mathbb{Z}_p$ is a UFD (unique factorization domain). The only irreducible (prime) element, up to associates, is p.*

That is, every nonzero element $\alpha \in \mathbb{Z}_p$ may be uniquely expressed as $\alpha = p^m \varepsilon$ where $m \in \mathbb{Z}$, $m \geq 0$ and $\varepsilon \in U(\mathbb{Z}_p)$.

Every rational number $r = b/a$ with $a, b \in \mathbb{Z}$ and $p \nmid a$ is also in $\mathbb{Z}_p$, since both a and b are, and a is a p-adic unit. We have $b/a = \{x_k\}$ where $ax_k \equiv b \pmod{p^k}$ for $k \geq 1$. The rational numbers r which have this form are those for which $\text{ord}_p(r) \geq 0$, since $\text{ord}_p(b/a) = \text{ord}_p(b) - \text{ord}_p(a)$. These are called *$p$-integral* rational numbers. Define

$$R_p = \left\{ \frac{n}{d} \in \mathbb{Q} : p \nmid d \right\} = \{x \in \mathbb{Q} \mid \text{ord}_p(x) \geq 0\}.$$

The set R_p of p-integral rationals is a subring both of $\mathbb{Q}$ and of $\mathbb{Z}_p$. Within $\mathbb{Z}_p$ they may be recognized as the p-adic integers whose digit sequence is ultimately periodic (just as the rationals are the real numbers with ultimately periodic decimal expansions).

Proposition 5.3.5. *R_p is a ring, with $\mathbb{Z} \subset R_p \subset \mathbb{Q}$, and $\mathbb{Z} \subset R_p \subset \mathbb{Z}_p$. Also, $R_p = \mathbb{Z}_p \cap \mathbb{Q}$.*

Corollary 5.3.6. (a) *Every rational number is in $\mathbb{Z}_p$ for all but a finite number of primes p.*
(b) *$\bigcap_{p \in \mathbb{P}} R_p = \mathbb{Z}$.*

We now extend the function ord_p, which we have already defined on $\mathbb{Z}$ and on $\mathbb{Q}$, to $\mathbb{Z}_p$. Since the prime p is fixed we may sometimes write ord instead of ord_p.

Definition 5.3.7. *For nonzero $\alpha \in \mathbb{Z}_p$ we define* $\text{ord}_p(\alpha) = m$ *where m is the largest integer for which $p^m|\alpha$ (in $\mathbb{Z}_p$). We also set* $\text{ord}_p(0) = \infty$.

So $\text{ord}_p(\alpha) = m \geq 0$ is the power of p appearing in its factorization $\alpha = p^m\varepsilon$. This definition agrees with the old definition of ord_p for rationals when $\alpha \in \mathbb{Z}_p \cap \mathbb{Q} = R_p$.

Proposition 5.3.8. *The function* $\text{ord}_p : \mathbb{Z}_p \to \mathbb{N}_0 \cup \{\infty\}$ *has the following properties:*
(1) *for $n \in \mathbb{Z}$ (or $\mathbb{Q}$), this definition of* $\text{ord}_p(n)$ *agrees with the one in Chapter 1;*
(2) $\text{ord}_p(\alpha\beta) = \text{ord}_p(\alpha) + \text{ord}_p(\beta)$*;*
(3) $\alpha|\beta \iff \text{ord}_p(\alpha) \leq \text{ord}_p(\beta)$*;*
(4) $\text{ord}_p(\alpha + \beta) \geq \min\{\text{ord}_p(\alpha), \text{ord}_p(\beta)\}$*, with equality if* $\text{ord}_p(\alpha) \neq \text{ord}_p(\beta)$.

We can also consider congruences in $\mathbb{Z}_p$. The next proposition shows that these are effectively the same as congruences in $\mathbb{Z}$ modulo powers of p.

Proposition 5.3.9. *For each $m \geq 0$, every $\alpha \in \mathbb{Z}_p$ is congruent modulo p^m to a unique integer n with $0 \leq n < p^m$. Moreover there is a ring isomorphism*

$$\mathbb{Z}_p/p^m\mathbb{Z}_p \cong \mathbb{Z}/p^m\mathbb{Z}.$$

5.4. **The field $\mathbb{Q}_p$.** Since the ring $\mathbb{Z}_p$ is an integral domain we can form its *field of fractions*, the field of *p-adic numbers* $\mathbb{Q}_p$:

$$\mathbb{Q}_p = \{\alpha/\beta \mid \alpha, \beta \in \mathbb{Z}_p, \beta \neq 0\}.$$

This forms a field under the usual rules for arithmetic of fractions, with $\mathbb{Z}_p$ as a subring and $\mathbb{Q}$ as a subfield. Since every nonzero p-adic integer has the form $p^n\varepsilon$ with ε a p-adic unit, we see that the nonzero elements of $\mathbb{Q}_p$ all have the form $x = p^m\varepsilon$ where now the exponent m is an arbitrary integer. We extend the order function from $\mathbb{Z}_p$ to a function $\text{ord}_p : \mathbb{Q}_p \to \mathbb{Z} \cup \{\infty\}$ by setting $\text{ord}_p(x) = m$. So $\mathbb{Z}_p = \{x \in \mathbb{Q}_p \mid \text{ord}_p(x) \geq 0\}$ (including 0 since $\text{ord}_p(0) = \infty$.) Parts (2) and (4) of Proposition 5.3.8 still apply.

$$\{0\} \subset \cdots \subset p^3\mathbb{Z}_p \subset p^2\mathbb{Z}_p \subset p\mathbb{Z}_p \subset \mathbb{Z}_p \subset p^{-1}\mathbb{Z}_p \subset p^{-2}\mathbb{Z}_p \subset p^{-3}\mathbb{Z}_p \cdots \subset \mathbb{Q}_p.$$

Let $x \in \mathbb{Q}_p \setminus \mathbb{Z}_p$, so $\text{ord}_p(x) = -m < 0$ and $x = p^{-m}\varepsilon$ with $\varepsilon \in U(\mathbb{Z}_p)$. Write $\varepsilon = a + p^m\beta$ with $\beta \in \mathbb{Z}_p$ and $a \in \mathbb{Z}$; by Proposition 5.3.9 this is uniquely possible with $0 \leq a < p^m$, and since ε is a unit, $p \nmid a$. Now

$$x = p^{-m}\varepsilon = p^{-m}(a + p^m\beta) = \frac{a}{p^m} + \beta;$$

so all p-adic numbers may be written (uniquely) as a p-adic integer plus a *fractional part* which is an ordinary rational number r satisfying $0 \leq r < 1$, with denominator a power of p.

Example: Let $x = \frac{1}{10} \in \mathbb{Q}_5$, with $\text{ord}_5(x) = -1$. Then $5x = \frac{1}{2} = 3 + 2 \cdot 5 + 2 \cdot 5^2 + 2 \cdot 5^3 + \ldots$ (using the method of earlier examples), so

$$x = 3 \cdot 5^{-1} + 2 + 2 \cdot 5 + 2 \cdot 5^2 + \ldots,$$

with fractional part $\frac{3}{5}$ and 5-integral part $x - \frac{3}{5} = -\frac{1}{2} = 2 + 2 \cdot 5 + 2 \cdot 5^2 + \ldots$.

Secondly, let $x = \frac{1}{100} \in \mathbb{Q}_5$, so $\text{ord}_5(x) = -2$ and $5^2x = \frac{1}{4} \in \mathbb{Z}_5$. To find the fractional part of x we approximate $\frac{1}{4}$ modulo 5^2 by solving $4y \equiv 1 \pmod{25}$ to get $y \equiv 19 \pmod{25}$. Then

$x - \frac{19}{25} = \frac{1-4\cdot 19}{100} = \frac{-75}{100} = -\frac{3}{4} \in \mathbb{Z}_5$, so the fractional part of x is $\frac{19}{25}$ and the 5-integral part is $-\frac{3}{4}$. (You can also get this by squaring $\frac{1}{10}$.)

We may use the ord_p function on $\mathbb{Q}_p$ to define a metric (distance function) and hence a topology on $\mathbb{Q}_p$. Then we may talk about convergence, continuity and such like; in particular, we will be able to justify the computations with infinite series we have seen in earlier examples. The key idea is that of a *norm* on a field.

Definition 5.4.1. *Let F be a field. A* norm *on F is a function $x \mapsto \|x\|$ from F to the real numbers satisfying the following properties:*

(i) *Positivity: $\|x\| \geq 0$, and $\|x\| = 0 \iff x = 0$;*
(ii) *Multiplicativity: $\|xy\| = \|x\|\,\|y\|$;*
(iii) *Triangle inequality: $\|x+y\| \leq \|x\| + \|y\|$.*

For example, the usual absolute value $|x|$ is a norm on the fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. We sometimes write this as $|x|_\infty$ by analogy with the p-adic norms introduced below. The *trivial norm*, defined by $\|x\| = 1$ for all nonzero x, is a norm on any field. Note that the multiplicativity and positivity always imply that $\|1\| = \|-1\| = 1$, so that $\|-x\| = \|x\|$ for all $x \in F$.

Given a norm $\|\cdot\|$ on F, we may use it to define a *metric* or distance function on F, by setting $d(x,y) = \|x-y\|$ for $x, y \in F$. This has the following properties:

(i) Positivity: $d(x,y) \geq 0$, and $d(x,y) = 0 \iff x = y$;
(ii) Symmetry: $d(x,y) = d(y,x)$;
(iii) Triangle inequality: $d(x,z) \leq d(x,y) + d(y,z)$.

The field F, equipped with the metric from a norm on F, becomes a metric space, and hence also a topological space, so that we may consider such concepts as convergence of sequences and continuous functions on F. If F has more than one norm, this will lead to different metrics and (in general) different topologies on F. However, if we just replace a norm $\|x\|$ by $\|x\|^\alpha$ for a positive real number α, then the metrics will be equivalent (in the sense of metric spaces) and the topologies the same. We call a pair of norms which are related in this way *equivalent*.

We now introduce the *p-adic norms* on the field $\mathbb{Q}$. Fix a prime number p. Recall that the function $\mathrm{ord}_p : \mathbb{Q} \to \mathbb{Z} \cup \{\infty\}$ has the following properties; these also hold in $\mathbb{Q}_p$.

Lemma 5.4.2. (1) *$\mathrm{ord}_p(xy) = \mathrm{ord}_p(x) + \mathrm{ord}_p(y)$;*
(2) *$\mathrm{ord}_p(x + y) \geq \min\{\mathrm{ord}_p(x), \mathrm{ord}_p(y)\}$, with equality if $\mathrm{ord}_p(x) \neq \mathrm{ord}_p(y)$.*

Definition 5.4.3. *Let p be a prime. For nonzero $x \in \mathbb{Q}_p$ we define the p-adic norm of x to be*

$$|x|_p = p^{-\mathrm{ord}_p(x)},$$

and set $|0|_p = 0$.

Proposition 5.4.4. *For each prime p the p-adic norm is a norm on $\mathbb{Q}$ and on $\mathbb{Q}_p$. It satisfies the following stronger form of the triangle inequality:*

$$|x + y|_p \leq \max\{|x|_p, |y|_p\}.$$

The associated p-adic metric $d(x, y) = |x - y|_p$ on $\mathbb{Q}_p$ satisfies

$$d(x, z) \leq \max\{d(x, y), d(y, z)\},$$

with equality if $d(x, y) \neq d(y, z)$.

A norm or metric which satisfies this stronger form of the triangle inequality is called *non-Archimedean*, in contrast to more familiar *Archimedean* metrics. This inequality is sometimes known as the "isosceles triangle principle", since it implies that in a space with a non-Archimedean metric every triangle is isosceles!.

Example: Consider the 5-adic norm on $\mathbb{Q}$. Take $x = \frac{3}{10}$ and $y = 40$. Since $\text{ord}_5(x) = -1$ and $\text{ord}_5(y) = 1$ we have $|x|_5 = 5$ and $|y|_5 = 5^{-1}$. The third side of the "triangle" with vertices 0, x, y has length $|x - y|_5$. Now $x - y = -\frac{397}{10}$ so $\text{ord}_5(x - y) = -1$, and hence $|x - y|_5 = 5 = |x|_5$.

Exercise: Prove the *Product Formula*: for every nonzero $x \in \mathbb{Q}$ we have

$$|x|_\infty \prod_{p \in \mathbb{P}} |x|_p = 1.$$

The main theorem on norms on the rational field $\mathbb{Q}$ states that (up to equivalence) the only norms are the ones we have seen:

Theorem 5.4.5. *[Ostrowski's Theorem] Every nontrivial norm on $\mathbb{Q}$ is equivalent either to the standard absolute value $|x|$ or to the p-adic norm $|x|_p$ for some prime p. All these norms are inequivalent.*

We omit the proof. The idea is that if $\|n\| \geq 1$ for all nonzero $n \in \mathbb{Z}$, then one can show that $\|x\| = |x|_\infty^\alpha$ for some $\alpha > 0$, while if $\|n\| < 1$ for some $n > 1$ then the least such n must be a prime p, and $\|x\| = \beta^{\text{ord}_p(x)}$ where $\beta = \|p\|$.

One can prove that $\mathbb{Q}_p$, with the p-adic metric, is complete. In fact, an alternative construction of $\mathbb{Q}_p$ is to start with the p-adic metric on $\mathbb{Q}$ and form the *completion* of $\mathbb{Q}$ with respect to this metric; this is entirely analogous to the construction of the real numbers by completing $\mathbb{Q}$ with respect to the usual metric. Either way we end up with a complete field $\mathbb{Q}_p$ in which $\mathbb{Q}$ is *dense* (we prove this below).

The theory of p-adic analysis has many counter-intuitive features, such as the fact that every p-adic triangle is isosceles. Another one is: a series $\sum_{n=1}^{\infty} a_n$ with terms $a_n \in \mathbb{Q}_p$ converges *if and only if* the terms tend to zero, i.e. $\lim_{n\to\infty} a_n = 0$. We will prove a special case of this in the next proposition.

Rather than continuing with this analytic theory, however, we will content ourselves with some examples, which in particular show that the earlier computations we carried out with power series are valid in $\mathbb{Q}_p$, once we have equipped it with its (p-adic) metric.

Proposition 5.4.6. (1) *Let $\alpha \in \mathbb{Z}_p$ be given by a coherent sequence $\{x_k\}$ of integers. Then $\lim_{k\to\infty} x_k = \alpha$, the limit being in the p-adic topology on $\mathbb{Z}_p$.*

(2) *Let $(a_i)_{i=0}^{\infty}$ be a sequence of integers with $0 \le a_i \le p-1$ for all $i \ge 0$. Then the series $\sum_{i=0}^{\infty} a_i p^i$ converges in $\mathbb{Z}_p$ to the p-adic integer $\alpha = \{x_k\}$, where $x_k = \sum_{i=0}^{k-1} a_i p^i$.*

Corollary 5.4.7. *Every p-adic integer in $\mathbb{Z}_p$ is the limit of a convergent sequence of rational integers. Every p-adic number in $\mathbb{Q}_p$ is the limit of a sequence of rational numbers.*

In other words, $\mathbb{Z}$ is *dense* in $\mathbb{Z}_p$, and $\mathbb{Q}$ is *dense* in $\mathbb{Q}_p$.

Examples:

$$\begin{aligned}
\sqrt{2} &= 3 + 1\cdot 7 + 2\cdot 7^2 + 6\cdot 7^3 + 7^4 + 2\cdot 7^5 + 7^6 + 2\cdot 7^7 + 4\cdot 7^8 + 6\cdot 7^9 + \cdots \in \mathbb{Z}_7;\\
40 &= 1 + 3 + 9 + 27 \in \mathbb{Z}_3 \text{ (a finite sum)};\\
-1 &= 2(1 + 3 + 3^2 + 3^3 + \dots) \in \mathbb{Z}_3;\\
-\frac{7}{8} &= 1 + 2\cdot 3 + 3^2 + 2\cdot 3^3 + 3^4 + \cdots \in \mathbb{Z}_3;\\
\frac{1}{3} &= 2 + 3\cdot 5 + 5^2 + 3\cdot 5^3 + 5^4 + 3\cdot 5^5 + 5^6 + \cdots \in \mathbb{Z}_5;\\
\frac{1}{10} &= 3\cdot 5^{-1} + 2 + 2\cdot 5 + 2\cdot 5^2 + \cdots \in \mathbb{Q}_5;
\end{aligned}$$

5.5. **Squares in $\mathbb{Z}_p$.** The method we used in Section 5.1 to find the 7-adic approximation to $\sqrt{2}$ is valid more generally. The case $p = 2$ is harder, so we start with odd primes.

Proposition 5.5.1. *Let p be an odd prime and $\alpha = \{x_k\} \in U(\mathbb{Z}_p)$. Then there exists $\beta \in \mathbb{Z}_p$ with $\alpha = \beta^2$ if and only if $\left(\frac{x_1}{p}\right) = +1$ (x_1 is a quadratic residue modulo p). In particular, every rational integer which is a quadratic residue modulo p is a p-adic square.*

An equivalent condition to $\left(\frac{x_1}{p}\right) = +1$ is $\left(\frac{a_0}{p}\right) = +1$ where a_0 is the first p-adic digit of α, since $\alpha \equiv x_1 \equiv a_0 \pmod{p}$. For $\alpha \in \mathbb{Z}_p$ we define $\left(\frac{\alpha}{p}\right) = \left(\frac{a_0}{p}\right) = \left(\frac{x_1}{p}\right)$.

Remark: A square unit in $\mathbb{Z}_p$ must have exactly two square roots, since $\mathbb{Z}_p$ is an integral domain, so the polynomial $x^2 - \alpha$ cannot have more than 2 roots. In the proof of the proposition one can see that after making an initial choice of y_1 as one of two possible choices for the square root modulo p, at all subsequent steps there is a unique choice.

An alternative approach to finding p-adic square roots is to start with a value $y = y_1$ which is a "first-order approximation", meaning a solution to $y^2 \equiv \alpha \pmod p$, and then iterate the map $y \mapsto y' = y + u(y^2 - \alpha)$ where u satisfies $1 + 2uy_1 \equiv 0 \pmod p$. At each step we obtain a better approximation, and in the limit we obtain an exact solution. To see why this works, the computation

$$(y')^2 - \alpha = (y + u(y^2 - \alpha))^2 - \alpha = (y^2 - \alpha)(1 + 2uy) + u^2(y^2 - \alpha)^2$$

shows that the valuation of $y^2 - \alpha$ strictly increases at each step, so $\beta = \lim y$ satisfies $\beta^2 - \alpha = \lim(y^2 - \alpha) = 0$.

Examples: 1. Taking $p = 7$ and $\alpha = 2$ we see that 2 is a 7-adic square since $\left(\frac{2}{7}\right) = 1$. One square root is $\beta = 3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + \ldots$ (see the calculation done in Section 5.1) and the other is $-\beta = 4 + 5 \cdot 7 + 4 \cdot 7^2 + 0 \cdot 7^3 \ldots$.

2. Take $p = 3$ and $\alpha = -2$. Using the second approach, take $y = 1$ which satisfies $y^2 \equiv -2 \pmod 3$ as a first approximation. Let $u = 1$ so that $1 + 2uy \equiv 0 \pmod 3$, and iterate $y \mapsto y + u(y^2 - \alpha) = y^2 + y + 2$. The first few values of y are (reducing the k'th one modulo 3^k):

$$1, 4, 22, 22, 22, 508, 508, 2695, \ldots.$$

Expanding 2695 to base 3 gives the expansion

$$\sqrt{-2} = 1 + 3 + 2 \cdot 3^2 + 2 \cdot 3^5 + 3^7 + \cdots \in \mathbb{Z}_3$$

where the next nonzero term is $a_{11} 3^{11}$ since $2695^2 + 2 = 3^{11} \cdot 41$, so $|\sqrt{-2} - 2695|_3 = 3^{-11}$. (The last statement should be checked carefully.)

Now we have identified the p-adic units which are squares, it is a simple matter to determine all the squares in $\mathbb{Z}_p$.

Proposition 5.5.2. *Let p be an odd prime. Let $\alpha = p^m \varepsilon$ be a nonzero p-adic integer with $m =$ ord(α) and $\varepsilon \in U(\mathbb{Z}_p)$. Then α is a square in $\mathbb{Z}_p$ if and only if m is even and $\left(\frac{\varepsilon}{p}\right) = 1$.*

The case of 2-adic squares is a little different: for a 2-adic unit to be a square, it is not sufficient to be a square modulo 2 (which is true for all 2-adic units since they are all congruent to $1 \pmod{2}$); they must be congruent to 1 modulo 8. This is due to the fact that odd integer squares are all congruent to 1 modulo 8. The next result is that being congruent to $1 \pmod{8}$ is sufficient for a 2-adic unit to be a square in $\mathbb{Z}_2$.

Proposition 5.5.3. *A 2-adic unit α is a square in $\mathbb{Z}_2$ if and only if $\alpha \equiv 1 \pmod{8}$.*

The proof shows how to find a 2-adic square root in practice: start with $y = 1$ and repeatedly replace y by $y' = y + 2^{k-1}$ where $k =$ ord$_2(y^2 - \alpha)$.

Example: We compute $\sqrt{17}$ in $\mathbb{Z}_2$, which exists since $17 \equiv 1 \pmod{8}$.
Start with $y = 1$. Then $y^2 - 17 = -16 = -2^4$, so replace y by $y + 2^3 = 9$.
Now $y^2 - 17 = 9^2 - 17 = 64 = 2^6$, so replace y by $y + 2^5 = 41$.

Now $y^2 - 17 = 41^2 - 17 = 2^7 \cdot 13$, so replace y by $y + 2^6 = 105$.

Now $y^2 - 17 = 105^2 - 17 = 2^8 \cdot 43$, so replace y by $y + 2^7 = 233$; and so on.

Thus we obtain a sequence $1, 9, 41, 105, 233, \ldots$ converging to $\sqrt{17} \in \mathbb{Z}_2$, and $\sqrt{17} = 1 + 2^3 + 2^5 + 2^6 + 2^7 + \ldots$.

Similarly we may compute (approximations to) $\sqrt{-7}$ in $\mathbb{Z}_2$, to get

$$\sqrt{-7} = \lim\{1, 5, 21, 53, 181, \ldots\} = 1 + 2^2 + 2^4 + 2^5 + 2^7 + 2^{14} + \ldots$$

with digit sequence $1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, \ldots$. The long block of zero digits comes from the fact that $181^2 + 7 = 32768 = 2^{15}$, so 181 is a rather good approximation to $\sqrt{-7}$ in $\mathbb{Z}_2$. We have $\mathrm{ord}(\sqrt{-7} - 181) = 14$, so $|\sqrt{-7} - 181|_2 = 2^{-14}$.

5.6. **Hensel lifting.** The process we used in the previous section to find p-adic square roots for odd p involves going from a solution of a congruence modulo p^k to a solution modulo p^{k+1}. This process is called "Hensel lifting" after Kurt Hensel (1861–1941), the inventor of p-adic numbers. It is the p-adic equivalent of refining an approximate real solution to an equation to a more precise solution, correct to more decimal places.

We will prove a quite general result which generalises the p-adic square root procedure for odd primes p, and also shows why $p = 2$ was different. Formally, this Hensel lifting is very similar to the Newton-Raphson method for finding roots of equations over $\mathbb{R}$.

Theorem 5.6.1. *[Hensel Lifting Theorem] Let $f(X) \in \mathbb{Z}_p[X]$ be a polynomial, and let $x_1 \in \mathbb{Z}_p$ satisfy $f(x_1) \equiv 0 \pmod{p}$ and $f'(x_1) \not\equiv 0 \pmod{p}$. Then there exists a unique $x \in \mathbb{Z}_p$ such that $f(x) = 0$ and $x \equiv x_1 \pmod{p}$.*

Example: Let p be odd and $a \in \mathbb{Z}$ a quadratic residue modulo p. Then a is a p-adic square: just take $f(X) = X^2 - a$ in the theorem with x_1 a solution to $x^2 \equiv a \pmod{p}$. The derivative condition is that $f'(x_1) = 2x_1 \not\equiv 0 \pmod{p}$, which holds since $p \neq 2$.

Example: Let p be prime and take $f(X) = X^p - X$. We know from Fermat's Little Theorem that f has p roots modulo p, one in each residue class. Hensel's Theorem says that f has p roots in $\mathbb{Z}_p$ also. One of these is 0; the others are $(p-1)$'st roots of unity in $\mathbb{Z}_p$. One way of constructing these will be in the exercises.

Remark: In this proof we have $y \equiv -a/f'(x_1) \equiv -(f(x_n)/p^n)/f'(x_1) \pmod{p}$, so

$$x_{n+1} = x_n + p^n y \equiv x_n - f(x_n)/f'(x_n) \pmod{p^{n+1}}.$$

Thus, Hensel lifting consists of starting with a "seed" $x = x_1$ which must be a simple root of f $\pmod{p}$, and iterating the map

$$x \mapsto x - f(x)/f'(x),$$

just as in the classical Newton method. Every iteration gives one more p-adic "digit", and the sequence always converges! To use the iteration formula to go from a root modulo p^n to a root modulo p^{n+1}, you can compute the inverse u of $f'(x_1) \pmod{p}$ once and for all at the start, and simply iterate $x \mapsto x - uf(x)$, as in the next example.

Example: We'll compute an approximation to $\sqrt[3]{2} \in \mathbb{Q}_5$. An initial approximation is $x_1 = 3$, and since $3^3 \equiv 2 \pmod{25}$ we can also take $x_2 = 3$. Here $f(X) = X^3 - 2$, so $f'(X) = 3X^2$

and $f'(x_1) = 27 \equiv 2 \pmod 5$ with inverse $u = -2$, so the recurrence is $x \mapsto x + 2(x^3 - 2)$:

$$x_3 \equiv 3 + 2(27-2) \equiv 53 \pmod{5^3}; \qquad \text{now } 53^3 \equiv 127 \pmod{5^4} \implies$$
$$x_4 \equiv 53 + 2(127-2) \equiv 303 \pmod{5^4}; \qquad \text{now } 303^3 \equiv 2502 \pmod{5^5} \implies$$
$$x_5 \equiv 303 + 2(2502-2) \equiv 5305 \equiv 2178 \pmod{5^5}; \qquad \text{and so on.}$$

We have an approximation to $\sqrt[3]{2}$, good to five 5-adic "digits":

$$\sqrt[3]{2} = 3 + 2 \cdot 5^2 + 2 \cdot 5^3 + 3 \cdot 5^4 + \cdots \in \mathbb{Q}_5.$$

This statement is analogous to saying that

$$\sqrt[3]{2} = 1.259921 \cdots = 1 + 2 \cdot 10^{-1} + 5 \cdot 10^{-2} + 9 \cdot 10^{-3} + \cdots \in \mathbb{R}.$$