

**G13NUM: NUMBER THEORY
COURSE NOTES 2002**

J. E. CREMONA

CONTENTS

Introduction: What is Number Theory?	2
Basic Notation	3
1. Factorization	4
1.1. Divisibility in \mathbb{Z}	4
1.2. Greatest Common Divisors in \mathbb{Z}	5
1.3. The Euclidean Algorithm in \mathbb{Z}	5
1.4. Primes and unique factorization	6
1.5. Unique Factorization Domains	8
2. Congruences	13
2.1. Definition and Basic Properties	13
2.2. The structure of $\mathbb{Z}/m\mathbb{Z}$	14
2.3. Euler's, Fermat's and Wilson's Theorems	15
2.4. Some Applications	16
2.5. The Chinese Remainder Theorem or CRT	17
2.6. The structure of U_m	19
3. Quadratic Reciprocity	20
3.1. Quadratic Residues and Nonresidues	20
3.2. Legendre Symbols and Euler's Criterion	20
3.3. The Law of Quadratic Reciprocity	21
4. Diophantine Equations	24
4.1. Sums of two and four squares	24
4.2. Pythagorean Triples	25
4.3. Legendre's Equation	26
4.4. Fermat's Last Theorem	27
5. p -adic Numbers	29
5.1. Motivating examples	29
5.2. Definition of \mathbb{Z}_p	30
5.3. The ring \mathbb{Z}_p	31
5.4. The field \mathbb{Q}_p	33
5.5. Squares in \mathbb{Z}_p	36

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 Level C module G1CANT (Algebraic Number Theory). The geometric approach is pursued further in the Level C module G1CRPC (Rational Points on 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 $n = 4$ (and possibly also $n = 3$).
 - (3) $y^2 = x^3 + 17$ has exactly 8 integer solutions (x, y) , namely $x = -2, -1, 2, 4, 8, 43, 52$ and one further value which you can find for yourselves. (Proving that there are no more solutions is harder.)
 - (4) Every positive integer n can be written as a sum of four squares (allowing 0 as a square), for example

$$47 = 36 + 9 + 1 + 1$$

but not all may be written as a sum of 2 or 3 squares. Which? We will answer the 2- and 4-square problems later.

- 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^{14} .
 - (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 $F_5 = 641 \times 6700417$, $F_6 = 274177 \times 67280421310721$, $F_7 = 59649589127497217 \times 5704689200685129054721$, \dots , and no more prime values have been discovered in the sequence.
 - (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?
 - (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 or not 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: positive integers p 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. 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 (Easy facts about divisibility). For all $a, b, \dots \in \mathbb{Z}$:

- (1) $a|b \implies a|kb$ ($\forall k \in \mathbb{Z}$);
- (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 : $\pm 1|a$ ($\forall a \in \mathbb{Z}$), and $a|\pm 1 \iff a = \pm 1$;
- (8) Special properties of 0 : $a|0$ ($\forall a \in \mathbb{Z}$), and $0|a \iff a = 0$.

Proposition 1.2 (Division Algorithm in \mathbb{Z}). Let $a, b \in \mathbb{Z}$ with $a \neq 0$. There exist unique integers q, r such that

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

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 2.1 (4) we have $(a) = (b) \iff a = \pm b$.

Recall that 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: $I \triangleleft R$. 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 1.3. Every ideal in \mathbb{Z} is principal.

Definition 2. 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 1.4. *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 3. *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 1.5 (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 1.6 (Euler's Lemma). *If $a|bc$ and $\gcd(a, b) = 1$ then $a|c$.*

If a_1, a_2, \dots, a_n is any finite sequence of integers then we similarly find that the ideal they generate, $I = (a_1, a_2, \dots, a_n) = \{k_1 a_1 + k_2 a_2 + \dots + k_n a_n \mid k_1, k_2, \dots, k_n \in \mathbb{Z}\}$ is an ideal of \mathbb{Z} , hence $I = (d)$ for a unique $d \geq 0$, and we define $d = \gcd(a_1, a_2, \dots, a_n)$. We say that a_1, a_2, \dots, a_n are coprime if $\gcd(a_1, a_2, \dots, a_n) = 1$. 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 the integers 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, \dots 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 4. 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 1.7. 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 proof of uniqueness of prime factorization relies on it. (It is easy to see that composite numbers do not have this property.) More generally:

Corollary 1.8. *Let p be a prime and $a_1, a_2, \dots, a_n \in \mathbb{Z}$. Then*

$$p|a_1a_2 \dots a_n \implies p|a_i \quad \text{for some } i.$$

Theorem 1.9 (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 1.10. *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 $p_1 < p_2 < \dots < p_k$ 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 $e_p \geq 0$, and only a finite number of the exponents e_p are nonzero.

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

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

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

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

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

Proposition 1.13. *Let $m, n \in \mathbb{Z}$ be nonzero. Then $\text{ord}_p(mn) = \text{ord}_p(m) + \text{ord}_p(n)$.*

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 into primes, applying ord_q to both sides (where q is some fixed prime) and using the Proposition gives

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

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

Proposition 1.14. *Let $n \in \mathbb{Z}$ be nonzero. Then n is a perfect square if and only if $n > 0$ and $\text{ord}_p(n)$ is even for all primes p .*

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

Theorem 1.15 (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 1.9 (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 5. (a) *A nonzero ring R is an Integral Domain if, for $a, b \in R$,*

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

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

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

Examples:

- \mathbb{Z} is an ED with $\lambda(n) = |n|$: this is what Proposition 1.2 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 :

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

Theorem 1.16. *The ring $\mathbb{Z}[i]$ of Gaussian Integers is a Euclidean Domain.*

Lemma 1.17. *The norm function N on $\mathbb{Z}[i]$ has the following properties:*

- (1) *Multiplicativity:* for all $\alpha, \beta \in \mathbb{Z}[i]$, $N(\alpha\beta) = N(\alpha)N(\beta)$;
- (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 1.18. *Let R be a Euclidean Domain. Then R is a Principal Ideal Domain.*

In a PID we have gcds just as in \mathbb{Z} , and Bezout’s identity, and the Euclidean algorithm can be used to compute them. In general we do not have uniqueness of gcds, only uniqueness up to associates (multiplication by a unit). (In \mathbb{Z} we avoided this non-uniqueness by insisting that all gcds were non-negative.)

Definition 6. *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 1.19. *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 1.20. *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 gcds 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 7. *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 1.21. *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 1.22. *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 1.23. *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 8. *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 $1 \leq j \leq r$ such that $q_j = v_jp_j$ and $u = vv_1v_2 \dots v_r$.

Theorem 1.24. *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? If $\pi \in \mathbb{Z}[i]$ is prime then π divides some ordinary “rational” prime p , since if $n = N(\pi) = \pi\bar{\pi}$ then $\pi|n$ so by primality of π , π divides at least one prime factor p of n .

If $N(\pi) = p$ is prime, then π must be irreducible: for if $\pi = \alpha\beta$ then $p = N(\pi) = N(\alpha)N(\beta)$, so one of $N(\alpha)$, $N(\beta)$ must be 1 and then one of α , β is a unit. For example, $1 + i$, $2 + i$, $3 + 2i$, $4 + i$ are prime since their norms are 2, 5, 13, 17. We will prove later that every rational prime p of the form $4k + 1$ can be expressed as a sum of two squares: $p = a^2 + b^2$; setting $\pi = a + bi$ we then have $p = N(\pi) = N(\bar{\pi})$, so π and $\bar{\pi}$ are both Gaussian primes. (π and $\bar{\pi}$ are not associate: exercise.) 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 a prime q is then also 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 $N(\alpha) = N(\beta) = q$, contradicting the fact that q is not a sum of two squares.

We sum up this example as follows; we have not quite proved it, since we have not yet proved that all primes of the form $4k + 1$ are sums of two squares.

Theorem 1.25. *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 $\pm 1, \pm i$. Its primes are as follows (together with their associates):*

- (1) $1 + i$, of norm 2;
- (2) each rational prime p of the form $4k + 1$ factorizes in $\mathbb{Z}[i]$ as $p = \pi\bar{\pi}$ where π and $\bar{\pi}$ are non-associate Gaussian primes of norm p ; explicitly, $\pi = a + bi$ where $p = a^2 + b^2$;
- (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), $2000 = (1 + i)^8(1 + 2i)^3(1 - 2i)^3$. If you want to explore the arithmetic of $\mathbb{Z}[i]$ further, try the `GaussInt` package in Maple.

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 $N(\alpha) = \alpha\bar{\alpha}$, 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 $\mathbb{Z}[i]$.

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 $N(\alpha) = \alpha\bar{\alpha}$; 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 $x^3 + y^3 = z^3$, and we will return to it later.

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 $w^2 + w + 4 = 0$, 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

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 9. 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 :

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

Here m is called the modulus. If $m \nmid (a - b)$ 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. 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 2.2. 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 2.3. (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 $n \mid m$ then $a \equiv b \pmod{n}$.

Lemma 2.4. 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 $a \mid m$, then $x \equiv y \pmod{m/a}$.

Proposition 2.5. Let $a, b \in \mathbb{Z}$. The congruence $ax \equiv b \pmod{m}$ has a solution $x \in \mathbb{Z}$ if and only if $\gcd(a, m) \mid b$. 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 $d \mid b$ (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 2.6. Each integer a is congruent modulo m to exactly one integer in the set

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

Definition 10. 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 nonnegative 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 11. 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 $\varphi: \mathbb{N} \rightarrow \mathbb{N}$ 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, \dots for integers and \bar{a}, \bar{b}, \dots for their residues in $\mathbb{Z}/m\mathbb{Z}$. Then one has $\bar{a} = \bar{b}$ (in $\mathbb{Z}/m\mathbb{Z}$) iff $a \equiv b \pmod{m}$ (in \mathbb{Z}), and $\mathbb{Z}/m\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}, \dots, \overline{m-1}\}$. 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 2.7. 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 2.8.

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

Definition 12. 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

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

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	15
$\varphi(m)$	1	1	2	2	4	2	6	4	6	4	10	4	12	6	8

Proposition 2.9. (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 2.10. 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 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 2.11. (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 2.12. *In $\mathbb{F}_p[X]$ the polynomial $X^p - X$ factorizes as a product of p linear factors:*

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

Corollary 2.13 (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 2.14. *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 2.15. *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 (proved below), 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 2.16. *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 $p = \pi\bar{\pi}$ 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 2.17. (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 2.18. *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 $p \equiv 1 \pmod{q}$.*

By clever use of suitable polynomials (as with $f(n)$ above) one can show that there are infinitely many primes $p \equiv 1 \pmod{m}$ for any m .

2.5. The Chinese Remainder Theorem or CRT.

Proposition 2.19 (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*

$$(1) \quad \begin{aligned} x &\equiv a \pmod{m} \\ x &\equiv b \pmod{n} \end{aligned}$$

have a solution which is unique modulo mn .

More generally, if $d = \gcd(m, n)$ then the congruences (1) have a solution if and only if $a \equiv b \pmod{d}$, and the solution (when it exists) is unique modulo $\text{lcm}(m, n) = mn/d$.

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 $(a \pmod{m}, b \pmod{n})$ and single residue classes $(c \pmod{mn})$ when m, n are coprime. This bijection is in fact a ring isomorphism:

Theorem 2.20 (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, \dots, 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 2.21. *Let $m, n \in \mathbb{N}$ be coprime. Then $\varphi(mn) = \varphi(m)\varphi(n)$.*

Corollary 2.22. *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, $\varphi(168) = 168 \cdot \left(1 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{3}\right) \cdot \left(1 - \frac{1}{7}\right) = 168 \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{6}{7} = 48$.

(2). $\varphi(100) = \varphi(4)\varphi(25) = 2 \cdot 20 = 40$.

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

Proposition 2.23. *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:

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

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 $1 = 7a + 13b = 14 - 13$, then $(1, -1)$ maps to $1(-13) - 1(14) \equiv -27 \pmod{91}$.

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 is rather like taking successive decimal approximations to an ordinary equation, and we will come back to this at the end of the module. 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 the for prime p , the multiplicative group $U_p (= \mathbb{F}_p^*)$ is cyclic.

Theorem 2.24. *Let p be a prime. Then the group $U_p = \mathbb{F}_p^*$ is cyclic.*

Definition 13. *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 2.25. *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 2.26. *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 and 2^{e-2} .

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 reduced to the question of taking square roots.

3.1. Quadratic Residues and Nonresidues.

Definition 14. 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 $p \mid x^2 - a \implies p \mid x^2 - b^2 \implies p \mid (x - b)(x + b) \implies p \mid (x - b)$ or $p \mid (x + b)$.

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 $\pm 2, \pm 5, \pm 6$.

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 $0 \leq k \leq p-2$ 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 15. The Legendre Symbol $\left(\frac{a}{p}\right)$ 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 $1 + \left(\frac{a}{p}\right)$.

Proposition 3.1. 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. *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 $\pm 1 \pmod{8}$ ”:

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

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

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 important Law of Quadratic Reciprocity.

3.3. The Law of Quadratic Reciprocity.

Proposition 3.3 (Gauss's Lemma). *Let p be an odd prime and a an integer not divisible by p . Consider the least residues of the integers ka for $1 \leq k \leq (p-1)/2$, reduced to lie between $-p/2$ and $p/2$. If the number of these which are negative is s , then $\left(\frac{a}{p}\right) = (-1)^s$.*

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, these are the integers between 1 and 6 up to sign. 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, $6^2 = 36 \equiv 10 \pmod{13}$.

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

Proposition 3.4. *Let p be an odd prime. Then*

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

Since the Legendre symbol is completely multiplicative $\left(\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)\right)$, and we know the values of $\left(\frac{-1}{p}\right)$ and $\left(\frac{2}{p}\right)$ for all p , all that remains is to evaluate $\left(\frac{q}{p}\right)$ for odd primes q different from p . The Law of Quadratic Reciprocity says that $\left(\frac{q}{p}\right)$ is very closely related to $\left(\frac{p}{q}\right)$ – a fact which is far from obvious. Special cases of the reciprocity law were conjectured by Euler on the basis of substantial calculations and knowledge, but it was Gauss who first proved it, and in fact gave several proofs. Our proof will be based on Gauss’s Lemma and is of a geometrical nature.

Theorem 3.5 (Quadratic Reciprocity). *Let p and q be distinct odd primes. Then*

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

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 $p \equiv q \equiv 3 \pmod{4}$.

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

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

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 on a modulo $4p$ and sometimes only on a modulo p . Here is a precise statement of this.

Proposition 3.6. *Let p and q be odd primes and a an integer coprime to pq .*

- *If $p \equiv q \pmod{4a}$ then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$;*

- If $p \equiv -q \pmod{4a}$ and $a > 0$ then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.

Alternatively, it is not hard to deduce this Proposition from Gauss's Lemma, and then we may deduce QR from it, as follows.

If $p \equiv q \pmod{4}$ then write $p - q = 4a$; then we have $\left(\frac{q}{p}\right) = \left(\frac{p-4a}{p}\right) = \left(\frac{-4a}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{a}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{a}{q}\right) = \left(\frac{-1}{p}\right)\left(\frac{4a}{q}\right) = \left(\frac{-1}{p}\right)\left(\frac{p-q}{q}\right) = \left(\frac{-1}{p}\right)\left(\frac{p}{q}\right)$, which is $\left(\frac{p}{q}\right)$ if $p \equiv q \equiv 1 \pmod{4}$ and $-\left(\frac{p}{q}\right)$ if $p \equiv q \equiv 3 \pmod{4}$.

Similarly, if $p \equiv -q \pmod{4}$ then write $p + q = 4a$ with $a > 0$; then $\left(\frac{q}{p}\right) = \left(\frac{4a-p}{p}\right) = \left(\frac{4a}{p}\right) = \left(\frac{a}{p}\right) = \left(\frac{a}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{p+q}{q}\right) = \left(\frac{p}{q}\right)$.

4. DIOPHANTINE EQUATIONS

A Diophantine Equation is simply an equation in one or more variables for which **integer** 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 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.

4.1. Sums of two and four squares. We start by collecting together results proved earlier about which primes can be written as sums of two squares.

Proposition 4.1. *Let p be a prime. Then $p = x^2 + y^2$ has a solution with $x, y \in \mathbb{Z}$ if and only if $p = 2$ or $p \equiv 1 \pmod{4}$.*

Our aim is to determine exactly which positive integers n are sums of two squares. This will turn out to depend on the prime factors of n .

Lemma 4.2. *Let $n = x^2 + y^2$ and suppose that n has a prime divisor $q \equiv 3 \pmod{4}$. Then $q|x$ and $q|y$, so $q^2|n$ and $n/q^2 = (x/q)^2 + (y/q)^2$.*

By repeated use of this lemma, every positive integer n which is a sum of two squares must have the form $n = ab^2$ where a is also a sum of two squares and has no prime factors congruent to $3 \pmod{4}$. This condition is in fact also sufficient.

Theorem 4.3. *The positive integer n may be expressed as a sum of two squares, $n = x^2 + y^2$, if and only if $\text{ord}_q(n)$ 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: We can interpret this result as a statement about norms of Gaussian integers, since $n = x^2 + y^2 \iff n = N(\alpha)$ where $\alpha = x + iy \in \mathbb{Z}[i]$. The result then can be deduced from the unique factorization of Gaussian integers together with the characterization of Gaussian primes given in Section 1.

One can similarly characterize positive integers of the form $n = x^2 + 2y^2$ as those such that $\text{ord}_q(n)$ is even for all primes $q \equiv 5, 7 \pmod{8}$. Either a direct proof or one based on unique factorization in the ring $\mathbb{Z}[\sqrt{-2}]$ is possible. A similar result holds for $n = x^2 + 3y^2$ (though is slightly harder to prove). 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.

Other possible generalizations of the expression $x^2 + y^2$ are possible: sums of two higher powers, and sums of more than two squares. We now look at the latter question.

Lemma 4.4. *Let n be a positive integer with $n \equiv 7 \pmod{8}$. Then n cannot be expressed as a sum of three squares, nor can 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 where the result is simpler, and easier to prove.

Theorem 4.5 (Lagrange). *Every positive integer may be expressed as a sum of four squares.*

Not that 0 is allowed as one of the squares. The Theorem will follow immediately from the following two lemmas.

Lemma 4.6. *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*

$$c_1 = a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4$$

$$c_2 = a_1 b_2 - a_2 b_1 + a_3 b_4 - a_4 b_3$$

$$c_3 = a_1 b_3 - a_3 b_1 - a_2 b_4 + a_4 b_2$$

$$c_4 = a_1 b_4 - a_4 b_1 + a_2 b_3 - a_3 b_2.$$

Lemma 4.7. *Every prime number may be expressed as a sum of four squares.*

4.2. 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

$$(2) \quad x^2 + y^2 = z^2.$$

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 (2) 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 4.8. *Let u and v be positive coprime integers with $u \not\equiv v \pmod{2}$ and $u > v$. Set*

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

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 subsection. 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.

4.3. 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 (mod 4).

The general equation

$$(3) \quad ax^2 + by^2 = cz^2$$

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 (3) we will always mean a solution other than the trivial one $(x, y, z) = (0, 0, 0)$, and 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 (equivalently, abc is square-free).

If $d = \gcd(a, b) > 1$ then we have a one-one correspondence between solutions (x, y, z_1) to the equation $(a/d)x^2 + (b/d)y^2 = (cd)z_1^2$ and (3) by setting $z = dz_1$. A similar reduction is possible 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 steps we achieve $\gcd(a, b) = \gcd(a, c) = \gcd(b, c) = 1$.

If $d^2|a$ then we have a one-one correspondence between solutions (x, y_1, z_1) to the equation $(a/d^2)x_1^2 + by_1^2 = cz_1^2$ and (3) by setting $y = dy_1$ and $z = dz_1$. Similarly with square factors of b or c .

From now on we assume that a, b, c are pairwise coprime and square-free.

Theorem 4.9. *Let $a, b, c \in \mathbb{N}$ be pairwise coprime and square-free. Then a non-trivial solution to (3) exists if and only if each of the quadratic congruences*

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

has a solution.

Remark: When there is one solution then the complete set of solutions may be parametrized as with Pythagorean Triples, expressing each of x, y, z as a homogeneous quadratic in two parameters u, v .

Example: Consider the equation $5x^2 + 7y^2 = 13z^2$. Solutions exist since the congruences $x^2 \equiv -ab \equiv -35 \equiv 4 \pmod{13}$, $x^2 \equiv ac \equiv 65 \equiv 9 \pmod{7}$ and $x^2 \equiv bc \equiv 91 \equiv 1 \pmod{5}$ all have solutions. We compute $5x^2 + 7y^2 - 13z^2$ for all triples (x, y, z) with $0 \leq x \leq 9$, $0 \leq y \leq 8$, $0 \leq z \leq 5$ (there are 540 such triples); of these, 5 triples give $5x^2 + 7y^2 - 13z^2 = 455 = abc$, and 3 give $5x^2 + 7y^2 - 13z^2 = 0$. So we find three solutions in this region, namely $(1, 4, 3)$, $(3, 1, 2)$, $(6, 2, 4)$, of which the third is not primitive.

There are much faster ways of solving the equation in practice than just searching, but our proof does show that a search of approximately abc triples will certainly produce a solution.

4.4. 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

$$(4) \quad x^n + y^n = z^n \quad \text{with } n \geq 3.$$

Fermat claimed, in the famous marginal note to his edition of the works of Diophantus, that there are no solutions to (4). 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 4.10 (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 4.11 (Fermat's Last Theorem for exponent 4). *The equation $x^4 + y^4 = z^4$ has no solutions in positive integers.*

This will follow from the following stronger statement: $x^4 + y^4$ cannot be a square, let alone a 4th power:

Theorem 4.12. *The equation $x^4 + y^4 = z^2$ has no solutions in positive integers.*

Corollary 4.13. *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.

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 often 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, but not always, 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 the following 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 which has no limit in \mathbb{Q} but does have a limit $\sqrt{2} = \lim_{k \rightarrow \infty} x_k$ 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:

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

We now consider the quadratic congruences

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

for $k = 1, 2, 3, \dots$. When $k = 1$ we have 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 making the initial choice of $x_1 = 3$ instead of -3) we find, uniquely,

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

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 $x^2 \equiv 2$

(mod 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 $3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + \dots$ 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 form a ring \mathbb{Z}_p containing \mathbb{Z} (one such ring for each prime p). And in the ring \mathbb{Z}_7 of 7-adic integers, our sequence $3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + \dots$ does 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 ($a_0 = 3, a_1 = 1, a_2 = 2, a_3 = 6, \dots$).

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 from \mathbb{Q} by a process almost identical to the construction of the real numbers. \mathbb{R} is the completion of \mathbb{Q} , complete in the usual analytic sense that Cauchy sequences converge in \mathbb{R} . One can define the real numbers as (equivalence classes of) Cauchy sequences of rational numbers, and we will start by defining p -adic integers as equivalence classes of suitable sequences of ordinary integers.

5.2. Definition of \mathbb{Z}_p . We fix once and for all a prime number p which will remain the same throughout (except in the examples).

Definition 16. A p -adic integer is given by a sequence of integers $\{x_k\}_{k=1}^{\infty} = \{x_1, x_2, x_3, \dots\}$, satisfying the condition

$$(5) \quad x_{k+1} \equiv x_k \pmod{p^k} \quad \text{for all } k \geq 1,$$

with two sequences $\{x_k\}$ and $\{y_k\}$ determining the same p -adic integer iff

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

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

A sequence satisfying (5) will be called *coherent*. Thus, a p -adic integer is actually an equivalence class of coherent sequences of ordinary integers.

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

The representation of a p -adic integer $x = \{x_k\}$ will be called *reduced* if $0 \leq x_k < p^k$ for all $k \geq 1$. Clearly every p -adic integer has a unique reduced representation.

Examples: Take $p = 3$. We have

$$\begin{aligned} 40 &= \{40, 40, 40, 40, 40, \dots\} = \{1, 4, 13, 40, 40, \dots\} \in \mathbb{Z}_3; \\ -1 &= \{-1, -1, -1, -1, -1, \dots\} = \{2, 8, 26, 80, 242, \dots\} \in \mathbb{Z}_3; \end{aligned}$$

the second representation is reduced in each case: $x_k = 40$ for all $k \geq 4$ when $x = 40$, while $x_k = 3^k - 1$ for all $k \geq 1$ when $x = -1$. 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.

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

$$(6) \quad x_k = a_0 + a_1 \cdot p + a_2 \cdot p^2 + \dots + a_{k-1} \cdot p^{k-1}$$

with each ‘‘digit’’ $a_i \in \{0, 1, 2, \dots, p-1\}$. The coherency condition (5) implies that $x_1 = a_0$, $x_2 = a_0 + a_1p$, $x_3 = a_0 + a_1p + a_2p^2$, 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 $0 \leq a_i \leq p-1$, and conversely every such digit sequence determines a unique p -adic integer $\alpha = \{x_k\}$ via (6). In the examples, the 3-adic digits of $20 = 2 + 2 \cdot 3^2$ are $2, 0, 2, 0, 0, \dots$ (effectively a finite sequence), while those of -1 form the infinite recurring sequence $2, 2, 2, 2, 2, \dots$.

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 $k \geq 1$, but for the moment this infinite series should be regarded as just a formal expression or shorthand.

5.3. The ring \mathbb{Z}_p . We may add and multiply p -adic integers, giving \mathbb{Z}_p the structure of a commutative ring, with \mathbb{Z} as a subring. We merely set

$$\begin{aligned} \{x_k\} + \{y_k\} &= \{x_k + y_k\}; \\ \{x_k\} \cdot \{y_k\} &= \{x_k y_k\}. \end{aligned}$$

One must check that the sequences on the right are coherent (in the sense of (5)), 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.

As our first step towards determining the structure of \mathbb{Z}_p , we have the following.

Lemma 5.1. *\mathbb{Z}_p is an integral domain.*

We will now develop a complete factorization theory for \mathbb{Z}_p , which turns out to be extremely simple. The first step is to determine the units $U(\mathbb{Z}_p)$.

Proposition 5.2. *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:

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

More generally, every rational number b/a with $p \nmid a$ is 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 just the ones 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. We set

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

So 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).

Corollary 5.3. 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 5.4. (a) Every rational number is in \mathbb{Z}_p for all but a finite number of primes p .

(b) $\bigcap_{p \in \mathbb{P}} R_p = \mathbb{Z}$.

Next we will show that \mathbb{Z}_p is a UFD, and it has only one prime (up to associates), namely p itself.

Proposition 5.5. 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 $\varepsilon \in U(\mathbb{Z}_p)$.

Corollary 5.6. For $\alpha \in \mathbb{Z}_p$, we have $p \mid \alpha$ if and only if $\alpha \notin U(\mathbb{Z}_p)$.

Corollary 5.7. \mathbb{Z}_p is a UFD (unique factorization domain). The only irreducible (prime) element, up to associates, is p .

Thus every p -adic integer is either a unit or a multiple of p , but not both. We have

$$\dots \subset p^3 \mathbb{Z}_p \subset p^2 \mathbb{Z}_p \subset p \mathbb{Z}_p \subset \mathbb{Z}_p$$

with each inclusion strict, and $U(\mathbb{Z}_p) = \mathbb{Z}_p \setminus p \mathbb{Z}_p$.

Definition 17. For nonzero $\alpha \in \mathbb{Z}_p$ we define $\text{ord}(\alpha) = m$, where $m \geq 0$ is the power of p appearing in the factorization above. Thus m is the largest integer for which $\alpha \in p^m \mathbb{Z}_p$. We set $\text{ord}(0) = \infty$.

We may also write ord_p instead of ord . The following proposition shows that the new definition agrees with the old definition of ord_p for integers when $\alpha \in \mathbb{Z}$.

Proposition 5.8. *The function $\text{ord} : \mathbb{Z}_p \rightarrow \mathbb{N}_0 \cup \{\infty\}$ has the following properties:*

- (1) $\text{ord}(n) = \text{ord}_p(n)$ for $n \in \mathbb{Z}$;
- (2) $\text{ord}(\alpha\beta) = \text{ord}(\alpha) + \text{ord}(\beta)$;
- (3) $\alpha|\beta \iff \text{ord}(\alpha) \leq \text{ord}(\beta)$;
- (4) $\text{ord}(\alpha + \beta) \geq \min\{\text{ord}(\alpha), \text{ord}(\beta)\}$, with equality if $\text{ord}(\alpha) \neq \text{ord}(\beta)$.

The equality $\mathbb{Q} \cap \mathbb{Z}_p = R_p$ proved earlier may now be seen as follows. Let $r = a/b \in \mathbb{Q} \cap \mathbb{Z}_p$. Then $b = ar$, so $\text{ord}_p(b) = \text{ord}(b) = \text{ord}(a) + \text{ord}(r) \geq \text{ord}(a) = \text{ord}_p(a)$, so $\text{ord}_p(r) = \text{ord}_p(a) - \text{ord}_p(b) \geq 0$, and hence $r \in R_p$.

Proposition 5.9. *\mathbb{Z}_p is a PID (principal ideal domain). The only nonzero ideals are the principal ideals $(p^m) = p^m\mathbb{Z}_p$ for $m \geq 0$.*

So the factorization theory and ideal theory in \mathbb{Z}_p is very simple. We can also consider congruences in \mathbb{Z}_p ; the following proposition shows that these are effectively the same as congruences in \mathbb{Z} modulo powers of p .

Proposition 5.10. *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 $p^m\varepsilon$ with ε a p -adic unit, we find 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 $\text{ord} : \mathbb{Q}_p \rightarrow \mathbb{Z} \cup \{\infty\}$ by setting $\text{ord}(x) = m$. So $\mathbb{Z}_p = \{x \in \mathbb{Q}_p \mid \text{ord}(x) \geq 0\}$. (Recall that $\text{ord}(0) = \infty$.) Parts (2) and (4) of Proposition 5.8 still apply. Note that for nonzero $x \in \mathbb{Q}_p$, either $x \in \mathbb{Z}_p$ or $x^{-1} \in \mathbb{Z}_p$ (or both) depending on whether $\text{ord}(x) \geq 0$ or $\text{ord}(x) \leq 0$. For p -adic units ε with $\text{ord}(\varepsilon) = 0$ we have both $\varepsilon \in \mathbb{Z}_p$ and $\varepsilon^{-1} \in \mathbb{Z}_p$.

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

Let $x \in \mathbb{Q}_p$ with $\text{ord}(x) = -m < 0$ so $x \notin \mathbb{Z}_p$. Then $x = p^{-m}\varepsilon$ 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 5.10 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 non-integral p -adic numbers x may be written (uniquely) as a p -adic integer plus a *fractional part* which is an ordinary rational number r with denominator $p^{-\text{ord}(x)}$ satisfying $0 < r < 1$.

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

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

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

Secondly, let $x = \frac{1}{100} \in \mathbb{Q}_5$, so $\text{ord}(x) = -2$ 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 $-\frac{3}{4}$.

We may use the ord 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 18. 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 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 . Every nonzero rational number x may be written uniquely in the form $x = p^m a/b$ where $a, b, m \in \mathbb{Z}$, $b > 0$, $p \nmid ab$ and $\text{gcd}(a, b) = 1$. We then set $\text{ord}_p(x) = m$, and $\text{ord}_p(0) = \infty$. This agrees with the definition of $\text{ord}_p(n)$ in case $x = n \in \mathbb{Z}$, and in fact $\text{ord}_p(n/d) = \text{ord}_p(n) - \text{ord}_p(d)$ for any representation of the rational number $x = n/d$ as a fraction.

Lemma 5.11. The function $\text{ord} : \mathbb{Q} \rightarrow \mathbb{Z} \cup \{\infty\}$ has the following properties:

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

Definition 19. Let p be a prime. For nonzero $x \in \mathbb{Q}$ we define the p -adic norm of x to be

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

and set $|0|_p = 0$.

Proposition 5.12. For each prime p the p -adic norm is a norm on \mathbb{Q} . It satisfies the following stronger form of the triangle inequality:

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

The associated p -adic metric on \mathbb{Q} 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. The preceding inequality is sometimes known as the “isosceles triangle principle”, since it implies that every triangle is isosceles (in a space with a non-Archimedean metric).

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 5.13 (Ostrowski’s Theorem). *Every nontrivial norm on \mathbb{Q} is equivalent either to the standard absolute value $|x|$ or to the p -adic norm $|\cdot|_p$ 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 $\beta = \|p\|$.

Since the ord_p function on \mathbb{Q} extends to \mathbb{Q}_p , the associated norm and metric also extend to \mathbb{Q}_p . It turns out that \mathbb{Q}_p is complete (with the associated topology). 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.

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. $\lim_{n \rightarrow \infty} a_n = 0$. 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 5.14. (1) Let $\alpha \in \mathbb{Z}_p$ be given by a coherent sequence $\{x_k\}$ of integers. Then $\lim_{k \rightarrow \infty} x_k = \alpha$, 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 $0 \leq a_i \leq p-1$ for all $i \geq 0$. 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 5.15. Every p -adic integer in \mathbb{Z}_p is the limit of a convergent sequence of (ordinary) integers. Every p -adic number in \mathbb{Q}_p is the limit of a sequence of rational numbers.

This result is usually expressed by saying that \mathbb{Z} is dense in \mathbb{Z}_p , and \mathbb{Q} is dense in \mathbb{Q}_p . Note that the rational numbers which appear in the second part all have denominators which are a power of p .

Examples:

$$\sqrt{2} = 3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + 7^4 + 2 \cdot 7^5 + 7^6 + 2 \cdot 7^7 + 4 \cdot 7^8 + 6 \cdot 7^9 + \dots \in \mathbb{Z}_7;$$

$$40 = 1 + 3 + 9 + 27 \in \mathbb{Z}_3 \text{ (a finite sum);}$$

$$-1 = 2(1 + 3 + 3^2 + 3^3 + \dots) \in \mathbb{Z}_3;$$

$$\frac{1}{3} = 2 + 3 \cdot 5 + 5^2 + 3 \cdot 5^3 + 5^4 + 3 \cdot 5^5 + 5^6 + \dots \in \mathbb{Z}_5;$$

$$\frac{1}{10} = 3 \cdot 5^{-1} + 2 + 2 \cdot 5 + 2 \cdot 5^2 + \dots \in \mathbb{Q}_5;$$

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.

5.5. Squares in \mathbb{Z}_p .

Proposition 5.16. 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.

Note that 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 may set $\left(\frac{\alpha}{p}\right) = \left(\frac{x_1}{p}\right)$ where $\alpha \equiv x_1 \pmod{p}$.

Remark: The above proof shows that a square unit in \mathbb{Z}_p will have exactly two square roots, since after the choice of y_1 (where there are exactly two possible choices for the square root of x_1 modulo p), at all subsequent steps of the construction there is a unique choice for y_{k+1} given y_k . This is as it should be, since \mathbb{Z}_p is an integral domain, so the polynomial $x^2 - \alpha$ cannot have more than 2 roots.

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 $\beta = 3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + \dots$ (see the calculation done in subsection 5.1) and the other is $-\beta = 4 + 5 \cdot 7 + 4 \cdot 7^2 + 0 \cdot 7^3 + \dots$

2. Take $p = 3$ and $a = -2$. In the above proof take $x_k = -2$ for all k , so $a = 0$ at each step. Start with $y_1 = 1$ which satisfies $y_1^2 \equiv -2 \pmod{3}$. At the general step, given $y_k^2 = -2 + 3^k b$ we set $y_{k+1} = y_k + 3^k t$ where $b + 2ty_1 \equiv 0 \pmod{3}$, so $t \equiv b \pmod{3}$. Substituting $t = b$ gives $y_{k+1} \equiv y_k + 3^k b \equiv y_k^2 + y_k + 2 \pmod{p^{k+1}}$. The first few terms of the y_k sequence are:

$$1, 4, 22, 22, 22, 508, 508, 2695, 2695, 2695, \dots$$

so (expanding 2695 to base 3)

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

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.)

Once 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 5.17. *Let p be an odd prime. Let $\alpha = p^m \varepsilon$ be a nonzero p -adic integer with $m = \text{ord} \alpha$ 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 2-adic squares need to be treated separately: 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 (mod 2)); they must be congruent to 1 modulo 8. This is essentially due to the fact that odd integer squares are all congruent to 1 modulo 8. In the proof below, there is a twist compared to the case of odd primes. Something must happen to allow for the fact that 1 has 4 square roots modulo 8, namely ± 1 and ± 3 , but in the integral domain \mathbb{Z}_2 elements cannot have more than two square roots. In the inductive step below, when going from y_k to y_{k+1} , we find that either there are two choices or none – but if none, then switching to the alternative choice at the previous step works.

Proposition 5.18. *Let α be a 2-adic unit. Then α is a square in \mathbb{Z}_2 if and only if $\alpha \equiv 1 \pmod{8}$.*

Despite the last remark, to find a 2-adic square root in practice we do the following: given y_k satisfying $y_k^2 \equiv a \pmod{2^k}$ we set $y_{k+1} = y_k$ if $y_k^2 \equiv a \pmod{2^{k+1}}$ and otherwise set $y_{k+1} = y_k + 2^{k-1}$.

Example: We compute $\sqrt{17}$ in \mathbb{Z}_2 , which exists since $17 \equiv 1 \pmod{8}$. At each stage we have two values of y_k , only one of which survives to become one of the values of y_{k+1} , and which differ by 2^{k-1} . The first two values are taken to be 1 and 5, so we are constructing the square root which is congruent to 1 (mod 4).

k	3	4	5	6	7	8	9	10
y_k	1	1	25	41	41	233	233	233
y'_k	5	9	9	9	105	105	489	745

For example, for $k = 7$ we have $y_7 = 41$ and $y'_7 = 105$, differing by $64 = 2^6$. Now $41^2 - 17 = 1664 = 2^7 \cdot 13$, so 41 is not one of the values of y_8 , but $105^2 - 17 = 11008 = 2^8 \cdot 43$, so 105 is one value of y_8 , the other being $105 + 2^7 = 233$.

Thus we obtain

$$\begin{aligned} \sqrt{17} &= \{1, 1, 1, 9, 9, 41, 105, 233, 233, 233, \dots\} \\ &= 1 + 2^3 + 2^5 + 2^6 + 2^7 + \dots \end{aligned}$$

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

$$\begin{aligned}\sqrt{-7} &= \{1, 1, 5, 5, 21, 53, 53, 181, 181, 181, 181, 181, 181, 181, \dots\} \\ &= 1 + 2^2 + 2^4 + 2^5 + 2^7 + 2^{14} + \dots,\end{aligned}$$

with digit sequence $1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, \dots$. 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}_p . We have $\text{ord}(\sqrt{-7} - 181) = 14$, so $|\sqrt{-7} - 181|_2 = 2^{-14}$.