MAT 214, FALL 2017

1. Prime Numbers

1.1. **Definitions.** What are the "simplest" numbers? We can all agree that 0 and 1 are very simple. Adding 0 to any number will not change that number, and multiplying 1 to any number will also not change it. So 1 is the simplest number with respect to multiplication.

The next simplest numbers, with respect to multiplication, are prime numbers. An amazing thing is that prime numbers "build up" all the integers. We now give two equivalent definitions of prime numbers.

Definition 1.1. *A* ***prime number*** p *is an integer bigger than* 1 *that has no proper divisors, namely the only positive divisors of* p *are* 1 *and* p*.*

Definition 1.2. *A* ***prime number*** p *is an integer bigger than* 1 *such that for any integers* a *and* b*, we have* $p|ab \Rightarrow p|a$ *or* $p|b$ *.*

Attention By definition, 1 is NOT a prime number. This fits our philosophy at the beginning: we want prime numbers to be precisely the "second simplest" kind of integers with respect to multiplication. Since 1 is the simplest, it should be the "second simplest".

Now let us give the first proof in this course.

Proof of Definition 1.2⇒Definition 1.1. Let $p \geq 1$ be any positive integer. Suppose d is a positive integer dividing p. In particular $d \leq p$. We want to prove that $d = 1$ or $d = p$.

Since $d|p$, there exists an integer m (automatically positive) such that $p = dm$. So $p|p = dm$. Definition 1.2 implies that either $p|d$ or $p|m$.

(case) If $p|d$, then $p \geq d$ since d is positive. But we already have $d \leq p$. So $d = p$.

(case) If $p|m$, then $p \leq m$ since m is positive. But $p = dm \geq m$ since $d \geq 1$. So $p = m$, and hence $d = 1$.

In summary, either $d = 1$ or $d = p$. So we are done. □

1.2. **gcd and Euclid's algorithm.** Fix some (nonzero) integer a. What numbers can be expressed as ax for some integer x? The answer is a tautology: all integers b such that $a|b$. We write the set of such numbers as $a\mathbb{Z}$.

Now suppose we fix two nonzero integers a and b. We use the following notation: Let $a\mathbb{Z}+b\mathbb{Z}$ be the set of all numbers which can be expressed as $ax + by$ for some $x, y \in \mathbb{Z}$. Then what is $a\mathbb{Z} + b\mathbb{Z}$?

Let us look at an example. Suppose $a = 35$ and $b = 13$. Dividing 35 by 13, we get

$$35 = 2 \cdot 13 + 9.$$

Therefore $35\mathbb{Z}+13\mathbb{Z}=(2\cdot 13+9)\mathbb{Z}+13\mathbb{Z}=13\mathbb{Z}+9\mathbb{Z}$. (Check this) Repeat this process:

$$\begin{aligned} 13&=1\cdot 9+4\Rightarrow 13\mathbb{Z}+9\mathbb{Z}=9\mathbb{Z}+4\mathbb{Z},\\ 9&=2\cdot 4+1\Rightarrow 9\mathbb{Z}+4\mathbb{Z}=4\mathbb{Z}+1\mathbb{Z},\\ 4&=4\cdot 1+0\Rightarrow 4\mathbb{Z}+1\mathbb{Z}=1\mathbb{Z}+0\mathbb{Z}=\mathbb{Z}. \end{aligned}$$

The last step can be omitted as $4\mathbb{Z}\subset 1\mathbb{Z}=\mathbb{Z}$, so $4\mathbb{Z}+1\mathbb{Z}=\mathbb{Z}$. To sum it up, we have $35\mathbb{Z}+13\mathbb{Z}=\mathbb{Z}$.

In this process, we only used the division algorithm for $\mathbb{Z}$. So this process works for any two positive integers.

Let us look at another example. Suppose $a=8$ and $b=6$. We have

$$\begin{aligned} 8&=1\cdot 6+2\Rightarrow 8\mathbb{Z}+6\mathbb{Z}=6\mathbb{Z}+2\mathbb{Z},\\ 6&=3\cdot 2+0\Rightarrow 6\mathbb{Z}+2\mathbb{Z}=2\mathbb{Z}+0\mathbb{Z}=2\mathbb{Z}. \end{aligned}$$

So $8\mathbb{Z}+6\mathbb{Z}=2\mathbb{Z}$.

The process above is called the **Euclid's Algorithm**: start with two positive integers $a>b$; divide a by b to get a remainder r; replace a and b by b and r and repeat. The process will stop within finitely many steps. In the end we will get a positive integer c such that $a\mathbb{Z}+b\mathbb{Z}=c\mathbb{Z}$. Moreover, c is uniquely determined by a and b, namely if a and b are fixed, then there is only a such c.

Definition 1.3. *Let a and b be positive integers. Define the greatest common divisor of a and b, denoted by $\gcd(a,b)$, to be the unique positive integer such that $a\mathbb{Z}+b\mathbb{Z}=\gcd(a,b)\mathbb{Z}$.*

Remark 1.4. $\gcd(a,b)$ *satisfies the following properties:*

(1) $\gcd(a,b)|a$ *and* $\gcd(a,b)|b$*;*
(2) If $d|a$ *and* $d|b$*, then* $d|\gcd(a,b)$*. In particular* $d\leq\gcd(a,b)$*.*

Proof. (1) We have $a\in a\mathbb{Z}+b\mathbb{Z}=\gcd(a,b)\mathbb{Z}$. So $a=\gcd(a,b)x$ for some $x\in\mathbb{Z}$. So $\gcd(a,b)|a$. Similarly $\gcd(a,b)|b$.
(2) We have $\gcd(a,b)\in\gcd(a,b)\mathbb{Z}=a\mathbb{Z}+b\mathbb{Z}$. So $\gcd(a,b)=ax+by$ for some $x,y\in\mathbb{Z}$. But $d|a$ and $d|b$, so $d|ax+by=\gcd(a,b)$. In particular $d\leq\gcd(a,b)$ since $\gcd(a,b)>0$. □

1.3. **Proof of Definition 1.1 implies Definition 1.2.** Let $p\geq 1$ be any positive integer. Suppose a, b integers such that $p|ab$ and $p\nmid a$. We need to prove $p|b$.

Since $p\nmid a$ and p has no positive divisors except for 1 and p, so $\gcd(p,a)=1$. Hence $p\mathbb{Z}+a\mathbb{Z}=\mathbb{Z}$. But $1\in\mathbb{Z}$, so $1=px+ay$ for some $x,y\in\mathbb{Z}$. Multiplying both sides by b, we get $b=pbx+aby$. So $p|pbx+(ab)y=b$.

1.4. **The infinitude of primes.**

Theorem 1.5 (Euclid)**.** *There are infinitely many prime numbers.*

Proof. Suppose for a contradiction that there are only finitely many prime numbers, say $p_1,\ldots,p_m$. Consider $n=p_1\cdots p_m+1$. Since $n>1$, it has a prime factor. So $p_i|n$ for some $i\in\{1,\ldots,m\}$. But then $p_i|n-p_1\cdots p_m=1$. Contradiction. □

Note that in the proof, we claimed that every integer bigger than 1 has a prime factor. However we did not give a rigorous proof of this claim. To write down a rigorous proof, we need to use the **Principle of Induction**: If we want to prove a statement $\mathcal{P}(n)$ for all integers $n \geq n_0$, it suffices to prove:

- (base step) Prove $\mathcal{P}(n_0)$ is true;
- (induction step) Assume $\mathcal{P}(n_0), \ldots, \mathcal{P}(n-1)$ are true, prove $\mathcal{P}(n)$ is true.

Now we prove this claim. In this case $\mathcal{P}(n)$ is n has a prime divisor and $n_0 = 2$.

(base step) $\mathcal{P}(2)$ is true: 2 has a prime factor as 2 itself is a prime number.

(induction step) Assume $\mathcal{P}(2), \ldots, \mathcal{P}(n-1)$ are true. If n is a prime number, then n itself is a prime factor of n. If n is not a prime number, then by Definition 1.1 there exists an integer d with $2 \leq d \leq n-1$ such that $d|n$. So $\mathcal{P}(d)$ implies that d has a prime factor p. But $d|n$, so n has a prime factor p.

1.5. **The Unique Factorization Theorem.** Now we are ready to explain that integers are "built up" by prime numbers.

Theorem 1.6 (Unique Factorization Theorem)**.** *Every integer bigger than* 1 *has a factorization into a product of prime numbers, unique up to reordering. Namely, for any integer $n \geq 2$, the following properties hold:*

existence *There exist prime numbers $p_1, \ldots p_m$ such that $n = p_1 \cdots p_m$.*

(uniqueness) If $n = p_1 \cdots p_m = q_1 \cdots q_r$ with prime numbers $p_1 \leq \ldots p_m$ and $q_1 \leq \ldots q_r$, then $m = r$ and $p_i = q_i$ for any $i \in \{1, \ldots, m\}$.

Proof. (existence) We do induction on $n \geq 2$.

(base step) $\mathcal{P}(2)$ is true because 2 is a prime number.

(induction step) Assume $\mathcal{P}(2), \ldots, \mathcal{P}(n-1)$ are true.

If n is a prime number, then $n = n$ is a prime factorization.

If n is not a prime number, then by Definition 1.1 there exists an integer d with $2 \leq d \leq n-1$ such that $d|n$. So $n = dk$ for some integer k, and moreover $2 \leq k \leq n-1$ since $2 \leq d \leq n-1$.

Now $\mathcal{P}(d)$ implies that $d = p_1 \cdots p_s$ for some prime numbers $p_1, \ldots, p_s$, and $\mathcal{P}(k)$ implies that $k = q_1 \cdots q_t$ for some prime numbers $q_1, \ldots, q_t$. Hence we can express n as a product of prime numbers $n = p_1 \cdots p_s \cdot q_1 \cdots q_t$. □

2. Fundamental Theorem of Arithmetic and Induction

2.1. **Factorization in primes (continued).** Today we continue to finish the proof of UFT (Theorem 1.6). We do the uniqueness part. But first, let us prove the following lemma.

Lemma 2.1. *Let p be a prime number. Suppose that p divides the product $a_1 \cdots a_n$ of integers with $n \geq 2$. Then p divides a_i for some $i \in \{1, \ldots, n\}$.*

Proof. We prove the lemma by induction on n.

(base step) $\mathcal{P}(2)$ is true by Definition 1.2.

(induction step) Assume $\mathcal{P}(2), \ldots, \mathcal{P}(n-1)$ are true. Suppose $p | a_1 \cdots a_n = a_1(a_2 \cdots a_n)$. Then by $\mathcal{P}(2)$, we have either $p|a_1$ or $p|a_2 \cdots a_n$. If $p|a_1$, then we are done. If $p|a_2 \cdots a_n$, then by $\mathcal{P}(n-1)$, we have that $p|a_i$ for some $i \in \{2, \ldots, n\}$. To sum it up, p divides a_i for some $i \in \{1, \ldots, n\}$. So we are done. □

Now we are ready to finish the proof of the uniqueness part of Theorem 1.6.

Proof of Theorem 1.6 continued. $\boxed{\textit{uniqueness}}$ We do induction on $n \geq 2$.

(base step) $\mathcal{P}(2)$ is true because 2 is a prime number.

(induction step) Assume $\mathcal{P}(2), \ldots, \mathcal{P}(n-1)$ are true. Suppose $n = p_1 \ldots p_m = q_1 \ldots q_r$ for prime numbers $p_1 \leq \ldots \leq p_m$ and $q_1 \leq \ldots \leq q_r$.

Now $p_1 | n = q_1 \cdots q_r$, so p_1 divides q_i for some $i \in \{1, \ldots, r\}$ by Lemma 2.1. Recall that p_1 is a prime number, so in particular $p_1 \geq 2$. But q_i is also a prime number, so it has only two positive divisors 1 and q_i. So $p_1 = q_i$. Similarly $q_1 = p_j$ for some $j \in \{1, \ldots, m\}$.

Now we have $p_1 = q_i \geq q_1 = p_j \geq p_1$, so we must have $p_1 = q_i = q_1 = p_j$. In particular $p_1 = q_1$. So we get from $p_1 \ldots p_m = q_1 \ldots q_r$ the following equality

$$p_2 \ldots p_m = q_2 \ldots q_r.$$

Call this product n'. Then $1 \leq n' < n$ because $n' = n/p_1$ and $p_1 \geq 2$. If $n' = 1$, then both products are empty and we can already conclude. If $n' \geq 2$, then we can apply $\mathcal{P}(n')$ and get

$$m - 1 = r - 1, \ p_i = q_i \text{ for any } i \in \{2, \ldots, m\}.$$

So we are done. □

Another way of stating the Unique Factorization Theorem (Theorem 1.6) is as follows: Let $n \geq 2$ be an integer. Then there exist a unique set of prime numbers $\{p_1, \ldots, p_m\}$ and a unique set of positive integers $\{\alpha_1, \ldots, \alpha_m\}$ such that $n = p_1^{\alpha_1} \cdots p_m^{\alpha_m}$.

Sometimes when we want to study two or more integers at the same time, it is more convenient to allow the power to be 0. For example if we want to find $\gcd(a, b)$ for two positive integers a and b, then we can first factorize a and b under UFT

$$a = p_1^{\alpha_1} \cdots p_m^{\alpha_m},$$
$$b = p_1^{\beta_1} \cdots p_m^{\beta_m},$$

where we now require the α_i's and β_i's to be non-negative (so allow 0). In this way the statement of the result is cleaner: we have $\gcd(a, b) = p_1^{\min(\alpha_1, \beta_1)} \cdots p_m^{\min(\alpha_m, \beta_m)}$.

We can also define a notion of the **least common multiple** of a and b, denoted by $\mathrm{lcm}(a, b)$, as the smallest positive integer that is divisible by both a and b. Then with the notation above, it is easy to see that $\mathrm{lcm}(a, b) = p_1^{\max(\alpha_1, \beta_1)} \cdots p_m^{\max(\alpha_m, \beta_m)}$.

The following corollary is then easy to deduce.

Corollary 2.2. *Let a and b be positive integers. Then $ab = \gcd(a,b)\mathrm{lcm}(a,b)$.*

2.2. **Sum of powers of integers.** The following formulae can be proven by induction: For any integer $n \geq 1$, we have

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2},$$
$$1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6},$$
$$1^3 + 2^3 + \ldots + n^3 = \frac{n^2(n+1)^2}{4}.$$

2.3. **Fibonacci numbers.** The Fibonacci sequence is defined by $f_1 = 1$, $f_2 = 1$ and $f_n = f_{n-2} + f_{n-1}$ for $n \geq 3$. The first few Fibonacci numbers are

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$$

The following statements are true for all integers $n \geq 1$.

(1) $\gcd(f_n, f_{n+1}) = 1$.
(2) $f_{n+1} < (\frac{7}{4})^n$.
(3) $f_n = \frac{a^n - b^n}{\sqrt{5}}$ where $a = \frac{1+\sqrt{5}}{2}$ and $b = \frac{1-\sqrt{5}}{2}$.
(4) $f_1^2 + f_2^2 + \ldots + f_n^2 = f_n f_{n+1}$.
(5) $f_1 + f_3 + f_5 + \ldots f_{2n-1} = f_{2n}$.

These statements can be proven by induction.

Attention Except for (1), it is NOT enough to just prove $\mathcal{P}(1)$ as the base step because the truth of $\mathcal{P}(1)$ alone does not imply the truth of $\mathcal{P}(2)$. In fact due to the nature of the definition of the Fibonacci sequence, we need both $\mathcal{P}(n-2)$ and $\mathcal{P}(n-1)$ to deduce $\mathcal{P}(n)$. This suggests that the base step should contain both $\mathcal{P}(1)$ and $\mathcal{P}(2)$.

3. Modular Arithmetic

3.1. **The ring $\mathbb{Z}/m\mathbb{Z}$.** As we saw in Euclid's Algorithm, in order to compute $\gcd(a,b)$ for two integers $a > b$, we could iterate the division algorithm: For $a = qb + r$ with $r \in \{0, 1, \ldots, b-1\}$, we have $\gcd(a,b) = \gcd(b,r)$. In other words, what is important in the process is the division algorithm by b and the integer $r \in \{0, \ldots, b-1\}$ such that $b|a-r$.

In Modular Arithmetic, we strip away factors of the modulus b and consider only the remainder.

Definition 3.1. *Fix an integer m. We say two integers a and b are **congruent modulo** m, and write $a \equiv b \pmod{m}$, if $m|a-b$.*

*If two integers are congruent modulo m, we also say that they are in the same **congruence class modulo** m. We use $\mathbb{Z}/m\mathbb{Z}$ to denote the set of congruence classes modulo m.*

For example 35 and 9 are congruent modulo 13. They are in the same congruence class modulo 13.

Since $m|x \Leftrightarrow -m|x$ for any integer x, we may and do assume $m \geq 1$ in the rest of the discussion. For any integer x, we denote by $[x]$ the congruence class modulo m in which x lies. In other words, $[x] \in \mathbb{Z}/m\mathbb{Z}$.

By the Division Algorithm, we have

$$\mathbb{Z}/m\mathbb{Z} = \{[0], [1], \ldots, [m-1]\} \quad \text{as sets.}$$

Now we introduce a binary operation on $\mathbb{Z}/m\mathbb{Z}$, which we call the addition. It is defined as follows: for any $[a], [b] \in \mathbb{Z}/m\mathbb{Z}$, define $[a] + [b] = [a+b]$.

Theorem 3.2. $(\mathbb{Z}/m\mathbb{Z}, +; [0])$ *forms an abelian group.*

Proof. It is easy to check the 3 conditions for abelian groups: $[0] + [a] = [a]$, $[a] + [b] = [b] + [a]$ and $[a] + [-a] = [0]$ for any $[a], [b] \in \mathbb{Z}/m\mathbb{Z}$. However before checking these rules, we need to check that the binary operation $+$ on $\mathbb{Z}/m\mathbb{Z}$ is well-defined. Namely we need to prove: if $[a] = [a']$ and $[b] = [b']$, then $[a+b] = [a'+b']$.

Let us prove this. If $[a] = [a']$, then $m|a-a'$. Similarly $m|b-b'$. So

$$m|(a-a') + (b-b') = (a+b) - (a'+b').$$

So $[a+b] = [a'+b']$. □

3.2. **The group $(\mathbb{Z}/m\mathbb{Z})^\times$.** We can also define another binary operation $\cdot$ on $\mathbb{Z}/m\mathbb{Z}$ by letting $[a] \cdot [b] = [ab]$. As before, we can check that it is well-defined, namely if $[a] = [a']$ and $[b] = [b']$, then $[ab] = [a'b']$.

It is easy to see that the unit for $\cdot$ is $[1]$, namely $[1]$ is the only element of $\mathbb{Z}/m\mathbb{Z}$ such that $[1] \cdot [a] = [a]$ for any $[a] \in \mathbb{Z}/m\mathbb{Z}$. But what about taking the inverse with respect to $[1]$? Given $[a] \in \mathbb{Z}/m\mathbb{Z}$, is it possible to find an element $[b] \in \mathbb{Z}/m\mathbb{Z}$ such that $[a] \cdot [b] = [1]$?

If we do not modulo m, then the answer is rather easy. In this situation we are just considering $\mathbb{Z}$. The only integers which are invertible with respect to the multiplication is 1 or -1. For example $2 \cdot \frac{1}{2} = 1$, but $\frac{1}{2} \notin \mathbb{Z}$.

However the situation changes when we modulo m. Say $m = 5$, then $[2] \cdot [3] = [6] = [1]$ in $\mathbb{Z}/5\mathbb{Z}$. In other words, $[3] = [2]^{-1}$ in $\mathbb{Z}/5\mathbb{Z}$.

Let us make this discussion into the following lemma.

Lemma 3.3. *Let m and a be integers. Then*

$$\gcd(a,m)=1 \Leftrightarrow \textit{there exists an element } [b]\in\mathbb{Z}/m\mathbb{Z} \textit{ such that } [a]\cdot[b]=[1].$$

Proof. ($\Rightarrow$) We have $1\in\mathbb{Z}=a\mathbb{Z}+m\mathbb{Z}=\mathbb{Z}$ by assumption. So $1=ab+my$ for some $b,y\in\mathbb{Z}$. So in $\mathbb{Z}/m\mathbb{Z}$, we have

$$[1]=[ab]+[my]=[ab]=[a]\cdot[b].$$

($\Leftarrow$) Now $[ab]=[a]\cdot[b]=[1]$, so $m|ab-1$. So $ab-1=my$ for some $y\in\mathbb{Z}$. So $ab-my=1$. On the other hand $\gcd(a,m)\mathbb{Z}=a\mathbb{Z}+m\mathbb{Z}$. So $1\in\gcd(a,m)\mathbb{Z}$. So $\gcd(a,m)=1$. □

Now we make the following definition.

Definition 3.4. *Fix m an integer. We say that an element $[a]\in\mathbb{Z}/m\mathbb{Z}$ is **invertible** if it satisfies either of the two equivalent conditions in Lemma 3.3. The element $[b]$ is called the **inverse** of $[a]$ and be denoted by $[a]^{-1}$.*

We denote by $(\mathbb{Z}/m\mathbb{Z})^\times$ the set of invertible elements of $\mathbb{Z}/m\mathbb{Z}$.

Remark 3.5.
(1) Any invertible element $[a]$ has a unique inverse. In other words, if $[a][b]=[a][c]=[1]$, then $[b]=[c]$. The proof goes as follows: Multiplying the equality $[a][b]=[a][c]$ by $[b]$ (on the left), we get $([b][a])[b]=([b][a])[c]$. But $[b][a]=[a][b]=[1]$ by assumption, so we get $[b]=[c]$.
(2) For any $[a]\in(\mathbb{Z}/m\mathbb{Z})^\times$, we have $[a]^{-1}\in(\mathbb{Z}/m\mathbb{Z})^\times$. This is easy as $[a]$ is the inverse of $[a]^{-1}$ by definition.

Now we are ready to state the structural theorem.

Theorem 3.6. *$((\mathbb{Z}/m\mathbb{Z})^\times,\cdot;[1])$ is an abelian group.*

Proof. We need to show that $\cdot$ is indeed a binary operation on $(\mathbb{Z}/m\mathbb{Z})^\times$. Namely if $[a],[b]\in(\mathbb{Z}/m\mathbb{Z})^\times$, then $[a][b]\in(\mathbb{Z}/m\mathbb{Z})^\times$.

There are two ways to do this, each one using one equivalent definition of invertible elements.

(Method 1) Use the definition by gcd. Now $\gcd(a,m)=1$ by assumption, so $\gcd(ab,m)=\gcd(b,m)$ by the Exercise 1 of the first Assignment. But then $\gcd(ab,m)=1$ since $\gcd(b,m)=1$ by assumption. Hence we are done.

(Method 2) Use the abstract definition. We have

$$([a][b])([b]^{-1}[a]^{-1})=[a]([b][b]^{-1})[a]^{-1}=[a][1][a]^{-1}=[a][a]^{-1}=[1].$$

Hence we have found an inverse of $[a][b]$. So $[a][b]$ is invertible. □

What is $\#(\mathbb{Z}/m\mathbb{Z})^\times$? In other words, how many numbers among $1,2,\dots,m-1$ are coprime to m? This number is denote by $\phi(m)$ and is called the Euler-phi function of m. Next time we will give a formula for $\phi(m)$. Today let us look at an application.

Proposition 3.7. *If $ac\equiv bc \pmod m$ and $\gcd(c,m)=1$, then $a\equiv b \pmod m$.*

Proof. Let us consider $\mathbb{Z}/m\mathbb{Z}$. Then the assumptions are equivalent to:

(i) $[ac]=[bc]$;
(ii) $[c]$ is invertible, namely $[c]\in(\mathbb{Z}/m\mathbb{Z})^\times$.

By (ii), $[c]^{-1}$ exists in $\mathbb{Z}/m\mathbb{Z}$. Multiplying $[c]^{-1}$ on both sides of (i), we get

$$[ac][c]^{-1} = [bc][c]^{-1},$$

and hence

$$[a] = [b].$$

Thus $a \equiv b \pmod{m}$. □

Theorem 3.8. *If a is coprime to m, then $a^{\phi(m)} \equiv 1 \pmod{m}$.*

Proof. Consider $\mathbb{Z}/m\mathbb{Z}$. List the elements of $(\mathbb{Z}/m\mathbb{Z})^\times$ by $(\mathbb{Z}/m\mathbb{Z})^\times = \{[b_1], \ldots, [b_{\phi(m)}]\}$. We make the following claims.

(i) $[ab_i] \in (\mathbb{Z}/m\mathbb{Z})^\times$ for any $i \in \{1, \ldots, \phi(m)\}$.
(ii) If $[ab_i] = [ab_j]$, then $[b_i] = [b_j]$.

(i) is true because $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$ (by assumption) and the abelian group structure on $(\mathbb{Z}/m\mathbb{Z})^\times$. (ii) is true because $[ab_i] = [ab_j]$ implies $[a]^{-1}[ab_i] = [a]^{-1}[ab_j]$, and hence $[b_i] = [b_j]$.

By (ii), we have a set $\{[ab_1], \ldots, [ab_{\phi(m)}]\}$ which has $\phi(m)$ elements. By (i), we have

$$\{[ab_1], \ldots, [ab_{\phi(m)}]\} \subset (\mathbb{Z}/m\mathbb{Z})^\times.$$

But both sets has $\phi(m)$ elements, so they are equal. In other words we have given another listing of the elements of $(\mathbb{Z}/m\mathbb{Z})^\times$ by $(\mathbb{Z}/m\mathbb{Z})^\times = \{[ab_1], \ldots, [ab_{\phi(m)}]\}$.

Now take the product of all elements of $(\mathbb{Z}/m\mathbb{Z})^\times$. From the two listings of the elements of $(\mathbb{Z}/m\mathbb{Z})^\times$, we get

$$[b_1 \cdots b_{\phi(m)}] = [a]^{\phi(m)}[b_1 \cdots b_{\phi(m)}].$$

Now multiplying both sides (on the right) by $[b_{\phi(m)}]^{-1} \cdots [b_1]^{-1}$, we get $[1] = [a]^{\phi(m)} = [a^{\phi(m)}]$. Hence we are done. □

As a corollary, we get Fermat's Little Theorem.

Corollary 3.9 (Fermat's Little Theorem)**.** *Let p be a prime number. Then $a^p \equiv a \pmod{p}$ for any integer a.*

4. Chinese Remainder Theorem

4.1. **Euler-phi function (continued).**

Theorem 4.1. *Let $m \geq 2$ be a positive integer. Suppose $m = p_1^{\alpha_1} \cdots p_n^{\alpha_n}$ under the Unique Factorization with $\alpha_i \geq 1$. Then*

$$\phi(m) = m(1 - \frac{1}{p_1}) \cdots (1 - \frac{1}{p_n}) = m \prod_{p|m,\ p\ prime} (1 - \frac{1}{p}).$$

Proof. We prove this theorem by induction on the sum $\alpha(m) := \alpha_1 + \ldots + \alpha_n$.
(base step) For m with $\alpha(m) = 1$. In this case m is a prime number, so $\phi(m) = m-1 = m(1-\frac{1}{m})$.
(induction step) Assume the equality for any positive integer m' with $\alpha(m') < \alpha(m)$.

Let p be a prime factor of m and write $m = pm'$. Then $\alpha(m') = \alpha(m) - 1$. So by induction hypothesis we have

$$\phi(m') = m' \prod_{q|m',\ q \text{ prime}} (1 - \frac{1}{q}).$$

We need to consider two cases: $p|m'$ and $p \nmid m'$.

(case) If $p|m'$, then m and m' have the same prime factors. Thus an integer x is coprime to m if and only if x is coprime to m'.

Given any integer x among $1, 2, \ldots, m-1$, we can divide x by m' and obtain $x = m'a + r$ with $r \in \{0, 1, \ldots, m'-1\}$. Then $a \in \{0, 1, \ldots, p-1\}$ because $1 \leq x \leq m-1$. So x is coprime to m if and only if r is coprime to m'.

So there are $\phi(m')$ choices of r so that r is coprime to m'. There are p possibilities for a. Hence we get

$$\phi(m) = p\phi(m').$$

So we get the desired equality by induction hypothesis.

(case) If $p \nmid m'$, then an integer x is coprime to m if and only if x is coprime to m' and $p \nmid x$. As above, we express any integer $x \in \{1, 2, \ldots, m-1\}$ as $m'a + r$ for $r \in \{0, 1, \ldots, m'-1\}$ and $a \in \{0, \ldots, p-1\}$. Then x is coprime to m' if and only if r is coprime to m'.

Now fix an r that is coprime to m', we need to count the number of possibilities for $a \in \{0, \ldots, p-1\}$ so that $m'a + r$ is not divisible by p. But $p|m'a + r$ if and only if $[m'a] = [-r]$ in $\mathbb{Z}/p\mathbb{Z}$. Since $p \nmid m'$, the congruence class $[m']$ is invertible in $\mathbb{Z}/p\mathbb{Z}$. So $[a] = [-r][m']^{-1}$ in $\mathbb{Z}/p\mathbb{Z}$. Since a is an integer between 0 and $p-1$, we see that there is a unique choice for $a \in \{0, \ldots, p-1\}$ such that $m'a + r$ is divisible by p. In other words, there are $p-1$ possibilities for $a \in \{0, \ldots, p-1\}$ so that $m'a + r$ is not divisible by p.

So we have

$$\phi(m) = (p-1)\phi(m').$$

Again the desired equality follows by induction hypothesis. □

4.2. **Chinese Remainder Theorem.**

Consider the following question: What are the last two digits of 123^{456}? The question is equivalent to asking for the reduction of 123^{456} modulo 100. Since $\gcd(123, 100) = 1$ and $\phi(100) = 40$, we know that $123^{440} = (123^{40})^{11} \equiv 1 \pmod{100}$ by Theorem 3.8. Hence

$$123^{456} \equiv 123^{16} \equiv 23^{16} \pmod{100}.$$

Then we have

$$23^2 = 529 \equiv 29 \pmod{100}$$
$$29^2 = 841 \equiv 41 \pmod{100}$$
$$41^2 = 1681 \equiv 81 \pmod{100}$$
$$81^2 = 6561 \equiv 61 \pmod{100}.$$

So $123^{456} \equiv 61 \pmod{100}$. Thus the last two digits of 123^{456} is 61.

This is quite a lot of work computationally. Sometimes it will not work: finding the last two digits of 8765^{4321} requires computing 65^{4321} modulo 100. As 65 and 100 are not coprime, we cannot apply Theorem 3.8. The Chinese Remainder Theorem, stated below, will allow us to understand reduction modulo 100 by understanding reduction modulo 4 and 25.

Theorem 4.2 (Chinese Remainder Theorem)**.** *Let a, b be coprime positive integers. For any integers x_1 and x_2, there exists a unique congruence class $[x]$ in $\mathbb{Z}/ab\mathbb{Z}$ such that*

$$x \equiv x_1 \pmod{a} \quad \text{and} \quad x \equiv x_2 \pmod{b}.$$

Proof. (existence) We consider integers x of the form $ay + x_1$. The second congruence then translates into $ay \equiv x_2 - x_1 \pmod{b}$. Since $\gcd(a, b) = 1$, such a y exists. Namely we can take y to be an integer among $0, 1, \ldots, b-1$ such that $[y] = [a]^{-1}[x_2 - x_1]$ in $\mathbb{Z}/b\mathbb{Z}$. Then this integer $ay + x_1$ satisfies the desired properties.

(uniqueness) Suppose x and x' are two integers satisfying the desired congruence conditions. Then $x - x'$ is divisible by both a and b. So $ab|x - x'$ since $\gcd(a, b) = 1$. So $[x] = [x']$ in $\mathbb{Z}/ab\mathbb{Z}$. □

To apply Theorem 4.2 to the above example, we write $100 = 4 \times 25$. Now $\gcd(4, 25) = 1$. We have $123 \equiv -1 \pmod{4}$ and so $123^{456} \equiv 1 \pmod{4}$. We also have $123 \equiv -2 \pmod{25}$ and $\phi(25) = 20$ and so $123^{456} \equiv (-2)^{16} = 65536 \equiv 11 \pmod{25}$. Hence the reduction of 123^{456} modulo 100 is congruent to 1 modulo 4 and to 11 modulo 25. Therefore it must be 61.

Let us now look at the example of 8765^{4321}. Again we have $8765 \equiv 1 \pmod{4}$ and so $8765^{4321} \equiv 1 \pmod{4}$. But $5|8765$, so $25 = 5^2|8765^2|8765^{4321}$. So $8765^{4321} \equiv 0 \pmod{25}$. Therefore $8765^{4321} \equiv 25 \pmod{100}$.

The last example will be a first step to solve Diophantine equation. Let us solve the congruence equation

$$x^3 + x \equiv 6 \pmod{12}. \tag{4.1}$$

That is, we want to find all the congruence classes in $\mathbb{Z}/12\mathbb{Z}$ whose 3-rd power plus itself is congruent to 6. It is of course possible to try all 12 congruence classes and see which ones work, but this is a lot of work, especially if 12 is replaced by a large integer. Instead we use the Chinese Remainder Theorem by first breaking up $12 = 3 \times 4$ in to a product of coprime integers. Now (4.1) is equivalent to the following two conditions combined:

$$x^3 + x \equiv 6 \equiv 0 \pmod{3}$$
$$x^3 + x \equiv 6 \equiv 2 \pmod{4}.$$

We can solve these two congruence equations by trying all congruence classes (the first one can also be solved by Fermat's Little Theorem). There are only 3 needed for the first one and 4

needed for the second one. Checking all of them we get

$$x \equiv 0 \pmod 3$$
$$x \equiv 1, 2, 3 \pmod 4.$$

There are then three solutions

$$\begin{array}{lllllll} x \equiv 0 & (\text{mod } 3), & x \equiv 1 & (\text{mod } 4) & \Rightarrow & x \equiv 9 & (\text{mod } 12) \\ x \equiv 0 & (\text{mod } 3), & x \equiv 2 & (\text{mod } 4) & \Rightarrow & x \equiv 6 & (\text{mod } 12) \\ x \equiv 0 & (\text{mod } 3), & x \equiv 3 & (\text{mod } 4) & \Rightarrow & x \equiv 3 & (\text{mod } 12) \end{array}$$

One can generalize the Chinese Remainder Theorem by including more congruence conditions. We delegate the proof to the homework.

Theorem 4.3. *Let $a_1, \ldots, a_n$ be pairwise coprime positive integers. For any integers $x_1, \ldots, x_n$, there exists a unique congruence class $[x]$ in $\mathbb{Z}/m\mathbb{Z}$, where $m = a_1 \cdots a_n$, such that $x \equiv x_i \pmod{a_i}$ for all $i \in \{1, \ldots, n\}$.*

The above example of solving congruence equations illustrate an important idea. In number theory, we usually have a polynomial $f(x, y, \ldots)$ with integer coefficients and we want to find integer solutions to the equation $f = 0$. For example Fermat's Last Theorem concerns with the polynomial $x^n + y^n - z^n$ and it states that there are no solutions other than $(0, 0, 0)$ when $n \geq 3$. Such a question is in general very hard while looking for solutions in $\mathbb{Z}/m\mathbb{Z}$ is much easier because $\mathbb{Z}/m\mathbb{Z}$ is finite. In this procedure we would want m to run over all positive integers. The Chinese Remainder Theorem (and its generalization) tells us that it suffices to consider only prime powers for m. As we will see later, for a fixed polynomial f, except for a finite set of primes, looking for solutions in $\mathbb{Z}/p^a\mathbb{Z}$ is the same as looking for solutions in $\mathbb{Z}/p\mathbb{Z}$. Finally since $\mathbb{Z}/p\mathbb{Z}$ is a field (every non-zero element is invertible) we will have more tools available to study the zeros of a polynomial in a field.

5. The finite field $\mathbb{F}_p$

5.1. **Wilson's Theorem.** Today we focus on the ring $\mathbb{Z}/m\mathbb{Z}$ where m is a prime number p. Recall that $(\mathbb{Z}/p\mathbb{Z})^\times = \{[1], \dots, [p-1]\}$ because $\gcd(a, p) = 1$ for any $a \in \{1, \dots, p-1\}$. In other words, every non-zero element of $\mathbb{Z}/p\mathbb{Z}$ is invertible.

Definition 5.1. *A **field** is a commutative ring $(F, +, \cdot; 0_F, 1_F)$ such that every non-zero element is invertible. In other words, $F^\times = \{a \in F : a \neq 0\}$. We sometimes abbreviate it as F if the other data ($+$, $\cdot$, 0_F, 1_F) are clear.*

Let us see some examples of fields. We have $(\mathbb{Z}/p\mathbb{Z}, +, \cdot; [0], [1])$. We sometimes denote this field by $\mathbb{F}_p$. The unit for addition is $[0]$, and the unit for multiplication is $[1]$.

Other examples include $\mathbb{Q}$ the field of rational numbers, $\mathbb{R}$ the field of real numbers, and $\mathbb{C}$ the field of complex numbers. Additions and multiplications are the usual ones in each case. The unit for addition is the number 0 and the unit for multiplication is the number 1 in each case.

In this list of examples, $\mathbb{Z}/p\mathbb{Z}$ is finite and $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are infinite.

Attention We only write $\mathbb{F}_p$ for $\mathbb{Z}/p\mathbb{Z}$. For an integer m which is not a prime number, we do NOT write $\mathbb{F}_m$ for $\mathbb{Z}/m\mathbb{Z}$. The reason will be explained in later lectures (in two or three weeks).

Recall the following Fermat's Little Theorem which we proved last Thursday.

Theorem 5.2 (Fermat's Little Theorem)**.** *$a^p = a$ for any $a \in \mathbb{F}_p$.*

In the statement, the symbol "a" represents an element of $\mathbb{F}_p$ instead of an integer. In other words, a means a congruence class mod p. This is why we don't write $[a]$.

Today we prove the following theorem.

Theorem 5.3 (Wilson's Theorem)**.** $(p-1)! \equiv -1 \pmod{p}$.

Proof. When $p = 2$ and $p = 3$, this is clearly true. From now on we assume $p \geq 5$.

Consider $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ and the list of congruence classes $[2], [3], \dots, [p-2]$. We observe that

$$\{[2], [3], \dots, [p-2]\} = \{[2]^{-1}, [3]^{-1}, \dots, [p-2]^{-1}\}.$$

This is because

(1) for any $a \in \{2, \dots, p-2\}$, $[a]^{-1}$ is again an invertible element of $\mathbb{Z}/p\mathbb{Z}$ and it is clearly not $[1]$ or $[p-1] = [-1]$;
(2) for any $a, b \in \{2, \dots, p-2\}$, $[a]^{-1} \neq [b]^{-1}$ if $a \neq b$.

Now we can break $\{[2], [3], \dots, [p-2]\}$, which has $p-3$ elements, into $(p-3)/2$ pairs consisting of a congruence class and its inverse. So we can regroup the product $[2][3]\cdots[p-2]$ so that

$$[2][3]\cdots[p-2] = [1]\cdots[1]((p-3)/2 \text{ of them}) = [1].$$

So $[1][2]\cdots[p-1] = [1]([2]\cdots[p-2])[p-1] = [1][1][p-1] = [-1]$ in $\mathbb{Z}/p\mathbb{Z}$. Hence we are done. □

Next time, we will see another proof of Wilson's Theorem using polynomials.

5.2. **The characteristic of a field.** Let $(F, +, \cdot; 0_F, 1_F)$ be a field. Let us define a map $\varphi\colon \mathbb{Z} \to F$, sending n to $n \cdot 1_F$ (meaning the sum of n-times 1_F). Using the fact $1_F^2 = 1_F$, it is not hard to show that $\varphi(n+m) = \varphi(n) + \varphi(m)$ and $\varphi(n \cdot m) = n \cdot \varphi(m)$. In the language of Tuesday's homework (if you have done it), φ is a ring homomorphism.

Definition 5.4. *The* ***characteristic*** *of the field F, denote by* $\mathrm{char}(F)$*, is the smallest positive integer n such that $\varphi(n) = 0_F$. In other words it is the smallest positive integer n such that $n \cdot 1_F = 0_F$. If no such integer exists, then we set* $\mathrm{char}(F) = 0$.

For example, $\mathrm{char}(\mathbb{F}_p) = p$ and $\mathrm{char}(\mathbb{Q}) = 0$.

Proposition 5.5. *We have the following properties.*

(1) $\mathrm{char}(F)$ *is either* 0 *or a prime number p;*
(2) for any $a \in F$, we have $\mathrm{char}(F) \cdot a = 0_F$*;*
(3) if $n \cdot 1_F = 0_F$ for some integer n, then $\mathrm{char}(F)$ *is a prime number and divides n.*

Proof. (1) Suppose for contradiction that $\mathrm{char}(F)$ is neither 0 nor a prime number. Then there exist integers $1 < d_1, d_2 < \mathrm{char}(F)$ such that $\mathrm{char}(F) = d_1 d_2$. However

$$0_F = (d_1 d_2) \cdot 1_F = (d_1 \cdot 1_F) \cdot (d_2 \cdot 1_F).$$

If $d_1 \cdot 1_F \neq 0_F$, then it is invertible in F since any non-zero element of F is invertible. But then multiplying both sides by $(d_1 \cdot 1_F)^{-1}$, we get $0_F = d_2 \cdot 1_F$. Thus either $d_1 \cdot 1_F = 0_F$ or $d_2 \cdot 1_F = 0_F$. This contradicts the minimality of $\mathrm{char}(F)$.

(2) We have

$$\mathrm{char}(F) \cdot a = \mathrm{char}(F) \cdot 1_F \cdot a = (\mathrm{char}(F) \cdot 1_F) \cdot a = 0_F \cdot a = 0_F.$$

(3) If $n \cdot 1_F = 0_F$ for some integer n, then $\mathrm{char}(F) \neq 0$. Hence $\mathrm{char}(F)$ is a prime number by (1). Recall that $\gcd(\mathrm{char}(F), n)\mathbb{Z} = \mathrm{char}(F)\mathbb{Z} + n\mathbb{Z}$, hence $\gcd(\mathrm{char}(F), n) = \mathrm{char}(F)x + ny$ for some $x, y \in \mathbb{Z}$. So

$$\gcd(\mathrm{char}(F), n) \cdot 1_F = (\mathrm{char}(F)x + ny) \cdot 1_F = x \cdot (\mathrm{char}(F) \cdot 1_F) + y \cdot (n \cdot 1_F) = 0_F.$$

By minimality of $\mathrm{char}(F)$, we have $\gcd(\mathrm{char}(F), n) = \mathrm{char}(F)$. So $\mathrm{char}(F) | n$.

□

Theorem 5.6. *(1) φ is injective if* $\mathrm{char}(F) = 0$*. In particular, if* $\mathrm{char}(F) = 0$*, then F is infinite.*
(2) φ induces an injection from $\mathbb{Z}/p\mathbb{Z}$ to F if $\mathrm{char}(F) = p$.

Proof. (1) If $\mathrm{char}(F) = 0$. Suppose $\varphi(n) = \varphi(m)$ for some integers n and m. Then

$$(n - m) \cdot 1_F = \varphi(n - m) = \varphi(n) - \varphi(m) = 0_F.$$

Because $\mathrm{char}(F) = 0$, we have $n - m = 0$ and so $n = m$. Thus φ is injective if $\mathrm{char}(F) = 0$.

But then F is infinite because $\mathbb{Z}$ is infinite.

(2) If $\mathrm{char}(F) = p$, then define a map $\overline{\varphi}\colon \mathbb{Z}/p\mathbb{Z} \to F$ by $\overline{\varphi}([a]) = \varphi(a)$. This map is well-defined: if $[n] = [m]$ in $\mathbb{Z}/p\mathbb{Z}$, then

$$\varphi(n) - \varphi(m) = \varphi(n - m) = \varphi(pr) = p \cdot \varphi(r) = 0_F,$$

and so $\varphi(n) = \varphi(m)$.

Let us prove that the map $\overline{\varphi}$ is injective. For $[m], [n] \in \mathbb{Z}/p\mathbb{Z}$, we have

$$\overline{\varphi}([m]) - \overline{\varphi}([n]) = \varphi(m) - \varphi(n) = (m - n) \cdot 1_F.$$

If $\overline{\varphi}([m]) - \overline{\varphi}([n]) = 0_F$, then $(m - n) \cdot 1_F = 0_F$. So $p | m - n$ by Proposition 5.5.(3), and hence $[m] = [n]$.

□

5.3. **A previous Exercise.** Ex 1 of the first Assignment. Show that if $\gcd(a, c) = 1$, then $\gcd(ab, c) = \gcd(b, c)$.

Proof. We shall use our definition of gcd, i.e. $\gcd(m, n)\mathbb{Z} = m\mathbb{Z} + n\mathbb{Z}$ and $\gcd(m, n) > 0$.

We have

$$a\mathbb{Z} + c\mathbb{Z} = \gcd(a, c)\mathbb{Z} = \mathbb{Z}.$$

Multiplying both sides with b, we get

$$ab\mathbb{Z} + bc\mathbb{Z} = b\mathbb{Z}.$$

Adding $c\mathbb{Z}$ on both sides, we have

$$ab\mathbb{Z} + (bc\mathbb{Z} + c\mathbb{Z}) = b\mathbb{Z} + c\mathbb{Z}.$$

But $bc\mathbb{Z} \subset c\mathbb{Z}$, so $bc\mathbb{Z} + c\mathbb{Z} = c\mathbb{Z}$. So we have

$$ab\mathbb{Z} + c\mathbb{Z} = b\mathbb{Z} + c\mathbb{Z}.$$

So $\gcd(ab, c)\mathbb{Z} = \gcd(b, c)\mathbb{Z}$. So $\gcd(ab, c) = \gcd(b, c)$ because both are positive. □

6. Polynomials with coefficients in a field

Through this section, we let $(F, +, \cdot; 0_F, 1_F)$ be a field. We abbreviate it as F.

Let $F[x]$ denote the set of polynomials with coefficients in F. We have

$$F[x] = \{a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 : n \geq 0,\ a_i \in F \text{ for all } i\}.$$

For example $[5]$, $[2]x^2 + [3]x$, $[5]x^3 + [6]x^2 + [4]$ are elements of $\mathbb{F}_7[x]$.

6.1. **Commutative ring structure.** In this course, we will see many analogies between $F[x]$ and $\mathbb{Z}$. Let us start with the following structural proposition: We can endow $F[x]$ with addition and multiplication. The formulae will be exactly as expected. We illustrate multiplication with an example: Suppose we are in $\mathbb{F}_7[x]$, then

$$([2]x^2 + [3]x)([4]x + [1]) = [2][4]x^3 + ([2][1] + [3][4])x^2 + [3][1]x = [1]x^3 + [3]x.$$

Then it is not hard to see that $0_F + f = f$ and $1_F \cdot f = f$ for any $f \in F[x]$. The following theorem is not hard to check.

Theorem 6.1. *$(F[x], +, \cdot; 0_F, 1_F)$ is a commutative ring.*

Next let us study the invertible elements of $F[x]$, namely the polynomials $f \in F[x]$ such that $f \cdot g = 1_F$ for some $g \in F[x]$. As for $\mathbb{Z}/m\mathbb{Z}$, we denote by $F[x]^\times$ the set of invertible elements of $F[x]$. We can prove that $(F[x]^\times, \cdot; 1_F)$ is an abelian group. The proof is similar to what we did for $(\mathbb{Z}/m\mathbb{Z})^\times$ (Method 2).

Recall that for $\mathbb{Z}/m\mathbb{Z}$, we had a concrete description of $(\mathbb{Z}/m\mathbb{Z})^\times$. We would like to ask whether such a concrete description exists for $F[x]^\times$. The answer is yes. Before computing this, let us look at the easy $\mathbb{Z}$.

Suppose now we want to compute $\mathbb{Z}^\times$. For any $a \in \mathbb{Z}^\times$, we have $ab = 1$ for some $b \in \mathbb{Z}$. Then we can take the absolute value of both sides and get $|a||b| = 1$. Note that $a, b \neq 0$, so $|a|, |b| \geq 1$. But then we must have $|a| = |b| = 1$. So $a = \pm 1$. On the other hand it is clear that $\pm 1 \in \mathbb{Z}^\times$. So $\mathbb{Z}^\times = \{\pm 1\}$.

In this easy computation, we used the absolute value of numbers to compare different numbers. If we can transport the notion to polynomials, then we will be able to compute $F[x]^\times$. This notion is the degree which we introduce now.

Definition 6.2. *To any polynomial $f \in F[x]$, we define its **degree**, denoted by $\deg(f)$, to be the largest integer n such that the coefficient of x^n is non-zero.*

In other words, for $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ with $a_n \neq 0_F$, we have $\deg(f) = n$. If $f = 0_F$, then we say $\deg(f) = -\infty$.

So in particular, if $f \neq 0_F$, then $\deg(f) \geq 0$. Moreover, the polynomials with degree 0 are precisely the non-zero constant polynomials.

Lemma 6.3. *For any $f, g \in F[x]$, we have $\deg(fg) = \deg(f) + \deg(g)$.*

Proof. If one of f and g is 0_F, then $fg = 0_F$. Hence $\deg(fg) = \deg(0_F) = -\infty$, and $\deg(f) + \deg(g) = -\infty$. So $\deg(fg) = \deg(f) + \deg(g)$ in this case.

Now suppose neither of f and g is 0_F. Suppose $\deg(f) = n$ and $\deg(g) = m$. Then we can write $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ with $a_n \neq 0_F$ and $g(x) = b_m x^m + b_{m-1}x^{m-1} + \ldots + b_0$ with $b_m \neq 0_F$. Thus

$$f(x)g(x) = a_n b_m x^{n+m} + \text{lower terms}.$$

Now $a_n, b_m \neq 0_F$ and F is a field, so $a_n b_m \neq 0_F$. So by definition $\deg(fg) = n + m$. So we are done. □

Now we are ready to compute $F[x]^\times$, just as we did for $\mathbb{Z}^\times$.

Theorem 6.4. *$F[x]^\times = F^\times$.*

Proof. Suppose $f \in F[x]^\times$, then $fg = 1_F$ for some $g \in F[x]$. Taking the degree of both sides, we get $\deg(f) + \deg(g) = \deg(fg) = \deg(1_F) = 0$. Thus neither of $\deg(f)$ and $\deg(g)$ is $-\infty$. But then $\deg(f), \deg(g) \geq 0$. So $\deg(f) = \deg(g) = 0$. So f is a non-zero constant polynomial. So $f \in F^\times = F \setminus \{0_F\}$. This proves $F[x]^\times \subset F^\times$.

On the other hand, every element of $F^\times$ is still invertible in $F[x]$. So to sum it up, we have $F[x]^\times = F^\times$. □

6.2. **Divisor and root.**

Definition 6.5. *We say that a polynomial $g(x)$ **divides** $f(x)$ in $F[x]$ if there exists a polynomial $h(x)$ such that $f(x) = g(x)h(x)$.*

For example in $\mathbb{F}_7[x]$, the polynomial $x - [3]$ divides $x^2 - [2]$ as

$$(x - [3])(x + [3]) = x^2 - [9] = x^2 - [2].$$

Note that any polynomial divides 0_F. If $g|f$ **and** $f \neq 0_F$, then $\deg(f) \geq \deg(g)$ (easy exercise).

Lemma 6.6. *Suppose $f \in F[x]$ satisfies $f(a) = 0_F$ for some $a \in F$. Then $x - a | f(x)$.*

Proof. The key trick is to use

$$x^r - a^r = (x - a)(x^{r-1} + x^{r-2}a + x^{r-3}a^2 + \ldots + a^{r-1}).$$

Suppose $f(x) = a_n x^n + \ldots + a_0$, then

$$\begin{aligned} f(x) &= f(x) - f(a) \\ &= a_n(x^n - a^n) + \ldots + a_1(x - a). \end{aligned}$$

Now $x - a$ divides each $x^r - a^r$ for $r \in \{1, \ldots, n\}$, so $x - a | f(x)$. □

Definition 6.7. *An element $a \in F$ such that $f(a) = 0_F$ is called a **root** of f.*

Proposition 6.8. *Let $f \in F[x]$ be a non-zero polynomial of degree d. Then f has at most d roots, with multiplicity counted.*

Proof. We prove by induction on d.

(base step) $\mathcal{P}(0)$ is true: when $d = 0$, f is non-zero constant, so it has no roots.

(induction step) Assume $\mathcal{P}(0), \ldots, \mathcal{P}(d-1)$ are true. Suppose f is a non-zero polynomial of degree d.

If f has no roots, then we are done. Otherwise suppose $f(a) = 0_F$. But $x - a | f$ by Lemma 6.6. So $f(x) = (x - a)^r g(x)$ for some $1 \leq r \leq d$ and some $g \in F[x]$.

If a' is a root of f other than a, then $0_F = f(a') = (a' - a)^r g(a')$, so $g(a') = 0_F$, namely a' is a root of g. So the roots of f are precisely a (with r times) and the roots of g.

Now $\deg(f) = \deg((x - a)^r) + \deg(g) = r + \deg(g)$. Now by $\mathcal{P}(d - r)$, g has at most $d - r$ roots. So f has at most $r + (d - r) = d$ roots. □

Now we use this to get another proof of Wilson's Theorem: $(p-1)! \equiv -1 \pmod{p}$ for any prime number p.

Another proof of Wilson's Theorem. Consider $f(x) = x^{p-1} - [1]$ in $\mathbb{F}_p[x]$. Fermat's Little Theorem implies that $[1], [2], \ldots, [p-1]$ are roots of f. So as in the previous proposition (this is sketchy here, think why), we get

$$f(x) = g(x)(x - [1]) \cdots (x - [p-1])$$

for some $g \in \mathbb{F}_p[x]$ by Lemma 6.6. Taking the degree of both sides, we get $p - 1 = \deg(f) = \deg(g) + p - 1$. So $\deg(g) = 0$. So $g(x)$ is a constant. Comparing the leading terms we get $g(x) = [1]$. Hence

$$x^{p-1} - [1] = (x - [1]) \cdots (x - [p-1]).$$

Now we can conclude by comparing the constant terms. □

7. Arithmetic in $F[x]$

7.1. **Irreducible polynomials.** Let us continue the analogy between $\mathbb{Z}$ and $F[x]$. Last time we introduced a notion of "size" which we call the degree and the notion of divisibility. Today we define "prime" elements of $F[x]$.

Definition 7.1. *An **irreducible polynomial** $f(x)$ is a non-constant polynomial that has no non-constant polynomial divisors of smaller degree, namely if $g|f$, then either g is constant or $\deg(g) = \deg(f)$.*

Definition 7.2. *An **irreducible polynomial** $f(x)$ is a non-constant polynomial such that $f|gh \Rightarrow f|g$ or $f|h$.*

Remark 7.3. *Let us see some easy consequences of the definitions.*

1. *The constant polynomial 0_F is NOT irreducible. Any non-zero constant polynomial is NOT irreducible. This is similar to $\mathbb{Z}$: 0 is not a prime number, and 1 (any element of $\mathbb{Z}^\times$) is also not a prime number.*
2. *Any polynomial of degree 1 is irreducible. This can be proven by Definition 7.1 and taking degrees.*
3. *If f is irreducible, then $cf(x)$ for any $c \in F^\times$ is again irreducible. This follows easily from Definition 7.1.*
4. *If $g|f$ and $\deg(g) = \deg(f)$, then $f(x) = cg(x)$ for some $c \in F^\times$. This can be proven by Definition 7.1 and taking degrees.*
5. *Recall that $\mathbb{Z}^\times = \{\pm 1\}$. So when defining prime numbers in $\mathbb{Z}$, we only considered positive numbers. Last time we saw $F[x]^\times = F^\times$, so when defining the "prime" elements of $F[x]$, we can confine ourselves to **monic** polynomials, i.e. polynomials whose leading coefficient is 1_F. This is justified by (3) and (4) of this remark. However we choose not to do this due to several reasons which will be revealed later in the course.*

Before going on, let us explain that the field F is important for this definition. Say $f(x) = x^2 + [1]$. If we view $f \in \mathbb{F}_2[x]$, then $f(x) = (x + [1])(x - [1])$ and so f is not irreducible. But if we view $f \in \mathbb{F}_3[x]$, then $f(x)$ is irreducible since it has no roots in $\mathbb{F}_3$.

As an exercise, please prove Definition 7.2⇒Definition 7.1 as we did for prime numbers in the first lecture. Proving Definition 7.1⇒Definition 7.2 requires the notion of gcd and Euclid's Algorithm. Recall that for $\mathbb{Z}$, the most important input for Euclid's Algorithm is the following division algorithm: suppose a, b are two positive integers, then there exist unique integers q and $r \in \{0, \dots, b-1\}$ such that $a = bq + r$. The analogous statement for $F[x]$ is as follows:

Proposition 7.4 (Division Algorithm)**.** *Let $a(x), b(x)$ be two polynomials in $F[x]$ and $b(x) \neq 0_F$. Then there exist unique polynomials $q(x)$ and $r(x)$ such that*

$$a(x) = b(x)q(x) + r(x),\ \deg(r) < \deg(b).$$

Proof. (uniqueness) Suppose there are polynomials $q_1(x), r_1(x), q_2(x), r_2(x)$ with

$$a(x) = b(x)q_1(x) + r_1(x),\ \deg(r_1) < \deg(b),$$
$$a(x) = b(x)q_2(x) + r_2(x),\ \deg(r_2) < \deg(b).$$

Subtracting these two equations gives $r_1(x) - r_2(x) = b(x)(q_2(x) - q_1(x))$. So $\deg(r_1 - r_2) = \deg(b) + \deg(q_2 - q_1)$. But $\deg(r_1 - r_2) \leq \max(\deg(r_1), \deg(r_2)) < \deg(b)$. So $r_1 - r_2 = 0_F$. So $r_1(x) = r_2(x)$ and hence $q_1(x) = q_2(x)$.

(existence) We do induction on $d = \deg(a) - \deg(b)$.
(base step) When $d < 0$, then $\mathcal{P}(d)$ holds: we can simply take $q(x) = 0_F$ and $r(x) = a(x)$.
(induction step) Now assume $d \geq 0$ and $\mathcal{P}(k)$ holds for any $k \leq d-1$. Note that $\deg(a) \geq 0$ now.

Put $m = \deg(a)$ and $n = \deg(b)$. Then $m - n = d \geq 0$. Write

$$a(x) = a_m x^m + \ldots + a_0,\ b(x) = b_n x^n + \ldots + b_0.$$

Multiplying $b(x)$ by $a_m b_n^{-1} x^{m-n}$ gives a polynomial of degree m with the same leading coefficient as $a(x)$. Now define $c(x) = a(x) - a_m b_n^{-1} x^{m-n} b(x)$, then $\deg(c) < m$. So $\deg(c) - \deg(b) < m - n = d$.

Now $\mathcal{P}(\deg(c) - \deg(b))$ implies that there exist polynomials $q'(x)$ and $r'(x)$ such that $c(x) = b(x)q'(x) + r'(x)$ with $\deg(r') < \deg(b)$. So we have

$$a(x) = b(x)(a_m b_n^{-1} x^{m-n} + q'(x)) + r'(x).$$

Now we can take $q(x) = a_m b_n^{-1} x^{m-n} + q'(x)$ and $r(x) = r'(x)$. □

Having the division algorithm, we can do what we did for $\mathbb{Z}$ to get Euclid's algorithm and find "greatest common divisors" of two polynomials. Before doing this, let us prove an auxiliary lemma.

Lemma 7.5. *Let $f, g \in F[x]$. Then $fF[x] = gF[x] \Leftrightarrow f(x) = cg(x)$ for some $c \in F^\times$.*

Proof. ($\Leftarrow$) For any $h \in F[x]$, we have $fh = g(ch) \in gF[x]$. So $fF[x] \subset gF[x]$. Conversely for any $h \in F[x]$, we have $gh = f(c^{-1}h) \in fF[x]$, so $gF[x] \subset fF[x]$. Hence $fF[x] = gF[x]$.

($\Rightarrow$) If one of f and g is 0_F, then the other one is also 0_F. In this case we are done.

Now assume neither of f and g is 0_F. By assumption we have $f \in fF[x] = gF[x]$. So $f = gh$ for some $h \in F[x]$. Taking the degrees we get $\deg(f) = \deg(g) + \deg(h)$. Since $f \neq 0_F$, we have $h \neq 0_F$. So $\deg(h) \geq 0$. So $\deg(f) \geq \deg(g)$.

Similarly $\deg(g) \geq \deg(f)$. So $\deg(f) = \deg(g)$. But then $\deg(h) = \deg(f) - \deg(g) = 0$. So h is a non-zero constant polynomial, i.e. $h \in F^\times$. □

Definition 7.6. *For two polynomials $f, g \in F[x]$, define $\gcd(f, g)$ to be a polynomial such that $fF[x] + gF[x] = \gcd(f, g)F[x]$.*

Remark 7.7.
(1) Under this definition, $\gcd(f, g)$ is NOT unique. But it is unique up to $F[x]^\times = F^\times$ (by Lemma 7.5), i.e. if two polynomials h_1 and h_2 satisfies $fF[x] + gF[x] = h_i F[x]$ for $i \in \{1, 2\}$, then $h_1 = ch_2$ for some $c \in F^\times$.
(2) As for $\mathbb{Z}$, we can prove: Let $h \in F[x]$. If $h|f$ and $h|g$, then $h|\gcd(f, g)$.

Now let us see an example of Euclid's algorithm. Say in $\mathbb{F}_7[x]$ we take (for simplicity we omit the "[]")

$$a(x) = x^3 - 2x^2 - 4x + 1,\ b(x) = 3x^2 - 4x - 4.$$

Now

$$x^3 - 2x^2 - 4x + 1 = (3x^2 - 4x - 4)(5x - 1) + (5x + 4)$$
$$3x^2 - 4x - 4 = (5x + 4)(2x - 1).$$

So $a(x)F[x] + b(x)F[x] = (5x + 4)F[x]$, or $\gcd(x^3 - 2x^2 - 4x + 1, 3x^2 - 4x - 4)$"="$5x + 4$.

Then with this we can prove Definition 7.1⇒Definition 7.2 as we did for $\mathbb{Z}$.

Again by imitating what we did for $\mathbb{Z}$, we can prove the following Unique Factorization Theorem for $F[x]$.

Theorem 7.8. *Every non-zero polynomial has a factorization into a product of a constant and monic irreducible polynomials, unique up to reordering. Namely $\forall f \in F[x]$ non-zero,*

- *there exist $c \in F^\times$ and monic irreducible polynomials $g_1, \ldots, g_n \in F[x]$ such that $f(x) = cg_1(x)\cdots g_n(x)$;*
- *if $f(x) = cg_1(x)\cdots g_n(x) = c'h_1(x)\cdots h_m(x)$ for $c, c' \in F^\times$ and $g_1, \ldots, g_n, h_1, \ldots, h_m$ monic irreducible polynomials, then $c = c'$, $n = m$ and $\{g_1, \ldots, g_n\} = \{h_1, \ldots, h_m\}$ as multi-sets (i.e. elements are counted with multiplicities).*

Again this theorem can be proven by induction on $\deg(f)$.

7.2. **Formal derivatives.**

Definition 7.9. *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \in F[x]$. Its **(formal) derivative**, denoted $f'(x)$, is*

$$f'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \ldots + a_1.$$

Here na means $a + a + \ldots + a$ the sum of n-times a.

For example $a'(x) = b(x)$ for $a(x) = x^3 - 2x^2 - 4x + 1, b(x) = 3x^2 - 4x - 4$ in $\mathbb{F}_7[x]$.

Proposition 7.10. *Let $f(x), g(x) \in F[x]$.*

(1) (Addition) $(f+g)' = f' + g'$.

(2) (Leibniz' rule) $(fg)' = f'g + fg'$.

Remark 7.11. *When $\mathrm{char}(F) = 0$, then $g' = 0_F \Rightarrow g$ is constant. However when $\mathrm{char}(F) = p$, then this is not true. Take for example $g(x) = x^p - 1$. Then $g' = px^{p-1} = 0_F$, but g is NOT constant.*

Proposition 7.12. *Let $f(x), g(x) \in F[x]$. Suppose $g(x)$ is irreducible and $g'(x) \neq 0_F$. Then $g(x)^2 | f(x) \Leftrightarrow g(x) | \gcd(f(x), f'(x))$.*

Proof. ($\Rightarrow$) We have $f(x) = g(x)^2 h(x)$ for some $h(x) \in F[x]$. So

$$\begin{aligned} f' &= (g^2)'h + g^2 h' \\ &= (g'g + gg')h + g^2 h' \\ &= g(2g'h + gh') \end{aligned}$$

So g divides both f and f'. So $g | \gcd(f(x), f'(x))$.

($\Leftarrow$) By assumption $g | f$. So $f = gh$ for some $h \in F[x]$. Then

$$f' = g'h + gh'.$$

Since $g | f' = g'h + gh'$, we have $g | g'h$. But g is irreducible, so $g | g'$ or $g | h$. But $g' \neq 0_F$ and $\deg(g') < \deg(g)$, so we cannot have $g | g'$. So $g | h$. Thus $g^2 | gh = f$. □

As we can see from the previous remark, the assumption $g' \neq 0_F$ is redundant if $\mathrm{char}(F) = 0$ (but should not be removed if $\mathrm{char}(F) = p$).

For example for $a(x) = x^3 - 2x^2 - 4x + 1$ in $\mathbb{F}_7[x]$, we can compute $a(x) = (x-2)^2(x+2)$. It is not so surprising now $\gcd(a, a') = 5(x-2)$.

8. Arithmetic in $F[x]$ (continued)

8.1. **Formal derivative (continued).** Let us prove the following lemma, which is part of Remark 7.11.

Lemma 8.1. *Assume* $\mathrm{char}(F) = 0$. *Let* $f \in F[x]$. *Then* $f' = 0_F \Leftrightarrow f$ *is constant.*

Proof. The implication $\Rightarrow$ is clear. Let us prove $\Leftarrow$ by contradiction.

Assume f is non-constant. Then we can write $f = a_n x^n + \ldots + a_0$ with $n \geq 1$ and $a_n \neq 0_F$. Then $f' = na_n x^{n-1} + \text{lower terms}$. Now $na_n = (n1_F)a_n$. If we could prove that $na_n \neq 0_F$, then $f' \neq 0_F$. This contradicts the assumption. So we are left to prove $na_n \neq 0_F$. But $a_n \neq 0_F$, so it suffices to prove $n1_F \neq 0_F$ because F is a field.

Since $n \neq 0$, we have $n1_F \neq 0_F$ because $\mathrm{char}(F) = 0$. So we are done. □

In this proof, we can see clearly where the assumption $\mathrm{char}(F) = 0$ is used. It is used in the last step. Suggested by its proof, we (guess and) prove its corresponding version for $\mathrm{char}(F) = p$.

Lemma 8.2. *Assume* $\mathrm{char}(F) = p$ *is a prime number. Let* $f \in F[x]$. *Then* $f' = 0_F \Leftrightarrow f = \sum_{i=0}^{n} a_i x^{ip}$ *for some* $a_i \in F$.

Proof. ($\Rightarrow$) We have

$$f' = \sum_{i=1}^{n} ipa_i x^{ip-1}.$$

Now $ipa_i = ip1_F a_i = (p1_F)(ia_i) = 0_F(ia_i) = 0_F$ for any $i \in \{1, \ldots, n\}$. So $f' = 0_F$.

($\Leftarrow$) Assume that there exists an integer $m \geq 1$ with $p \nmid m$ such that the m-th coefficient of f is non-zero. Then we can write $f = a_m x^m + \text{higher terms} + \text{lower terms}$ with $a_m \neq 0_F$. So

$$f' = ma_m x^{m-1} + \text{higher terms} + \text{lower terms}.$$

If we could prove $ma_m \neq 0_F$, then $f' \neq 0_F$. This contradicts the assumption. So we are left to prove $ma_m \neq 0_F$.

Assume again for contradiction that $ma_m = 0_F$. Then by Proposition 5.5.(3), we have $p = \mathrm{char}(F) | m$. But this contradicts our choice $p \nmid m$. Hence $ma_m \neq 0_F$. So we are done. □

Attention For $f \in F[x]$, the following fact is in general NOT true: $f = 0_F$ if $f(a) = 0_F$ for all $a \in F$. Take for example $F = \mathbb{F}_p$ and $f(x) = x^p - x$. By Fermat's Little Theorem, $f(a) = [0]$ for any $a \in \mathbb{F}_p$. But f is NOT the zero polynomial by definition. For example $\deg(f) = p \neq -\infty$.

As a consequence, for $f, g \in F[x]$, $f(a) = g(a)$ for all $a \in F \not\Rightarrow f = g$.

However, we will see in later lectures how to fix this problem (e.g. the statement is true if F is "large enough").

8.2. **Hensel's Lemma.**

Theorem 8.3 (Hensel's Lemma)**.** *Let* $f(x) \in \mathbb{Z}[x]$ *be a polynomial with integer coefficients. Suppose that* $x_1 \in \mathbb{Z}$ *satisfies*

$$f(x_1) \equiv 0 \pmod{p} \text{ and } f'(x_1) \not\equiv 0 \pmod{p}.$$

Then for any integer $r \geq 1$, *there exists an integer* x_r *such that*

$$x_r \equiv x_1 \pmod{p},\ f(x_r) \equiv 0 \pmod{p^r}.$$

Moreover, any two such x_r *are congruent to each other modulo* p^r.

Before proving Hensel's Lemma, let us first restate Hensel's Lemma in a form which better explains its meaning. To do this, we introduce the following notation.

Say $f \in \mathbb{Z}[x]$ and $m \geq 1$ is an integer. Then we use $(f \bmod m)$ to denote the polynomial in $(\mathbb{Z}/m\mathbb{Z})[x]$ which is obtained by modulo every coefficient of f by m. For example if $f = x^4 + 7x^3 + 5x^2 + 4x + 6$ and $m = 6$, then $(f \bmod 6) = x^4 + x^3 - x^2 - [2]x$ is a polynomial with coefficients in $\mathbb{Z}/6\mathbb{Z}$.

Now let $m \geq 1$ and $n \geq 1$ be two integers such that $m|n$. Then we can define a map $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$, sending $[a] \in \mathbb{Z}/n\mathbb{Z}$ to $[a] \in \mathbb{Z}/m\mathbb{Z}$. This map is well-defined: if $[a] = [a'] \in \mathbb{Z}/n\mathbb{Z}$, then $n|a - a'$, so $m|a - a'$, and hence a and a' are in the same congruence class modulo m. We call this map $(\bmod m)$.

Now we are ready to restate Hensel's Lemma.

Theorem 8.4. *Let $f(x) \in \mathbb{Z}[x]$ be a polynomial with integer coefficients. Suppose $x_1 \in \mathbb{Z}/p\mathbb{Z}$ satisfies*

$$(f \bmod p)(x_1) = [0] \text{ and } (f' \bmod p)(x_1) \neq [0] \text{ in } \mathbb{Z}/p\mathbb{Z}.$$

Then for any integer $r \geq 1$, there exists a unique $x_r \in \mathbb{Z}/p^r\mathbb{Z}$ such that

$$(\bmod p)(x_r) = x_1,\ (f \bmod p^r)(x_r) = [0] \text{ in } \mathbb{Z}/p^r\mathbb{Z}.$$

Now we can better understand the statement of Hensel's Lemma. The hypothesis in Hensel's Lemma is equivalent to: after modulo p, $x - x_1|(f \bmod p)$ but $(x - x_1)^2 \nmid (f \bmod p)$. We call such an x_1 a **simple root** of $(f \bmod p)$ (or of f over $\mathbb{F}_p$). Then the conclusion of Hensel's Lemma says that every simple root of f over $\mathbb{F}_p$ has a unique lift to a root of f in $\mathbb{Z}/p^r\mathbb{Z}$ (for any integer $r \geq 1$).

The following paragraph was said at the end of Section 4: In number theory, we usually have a polynomial $f(x, y, \ldots)$ with integer coefficients and we want to find integer solutions to the equation $f = 0$. For example Fermat's Last Theorem concerns with the polynomial $x^n + y^n - z^n$ and it states that there are no solutions other than $(0, 0, 0)$ when $n \geq 3$. Such a question is in general very hard while looking for solutions in $\mathbb{Z}/m\mathbb{Z}$ is much easier because $\mathbb{Z}/m\mathbb{Z}$ is finite. In this procedure we would want m to run over all positive integers. The Chinese Remainder Theorem (and its generalization) tells us that it suffices to consider only prime powers for m.

Now Hensel's Lemma tells us that for a fixed polynomial f, except for a finite set of primes, looking for solutions in $\mathbb{Z}/p^r\mathbb{Z}$ is the same as looking for solutions in $\mathbb{Z}/p\mathbb{Z}$. Since $\mathbb{Z}/p\mathbb{Z}$ is a field (every non-zero element is invertible) we will have more tools to study the zeros of a polynomial with coefficients in $\mathbb{Z}/p\mathbb{Z}$ compared to $\mathbb{Z}/p^r\mathbb{Z}$ $(r \geq 2)$.

Proof of Hensel's Lemma in the form of Theorem 8.3. We prove it by induction on $r \geq 1$. We will prove the existence part and uniqueness part at the same time, so the statement $\mathcal{P}(r)$ contains both parts.

(base step) The existence part of $\mathcal{P}(1)$ is true by assumption. The uniqueness part of $\mathcal{P}(1)$ is clearly true.

(induction step) Assume $\mathcal{P}(1), \ldots, \mathcal{P}(r-1)$ are true.
(existence part) By the existence part of $\mathcal{P}(r-1)$, there exists $x_{r-1} \in \mathbb{Z}$ such that $x_{r-1} \equiv x_1 \pmod p$ and $f(x_{r-1}) \equiv 0 \pmod{p^{r-1}}$. It suffices to find an integer a such that $f(x_{r-1} + p^{r-1}a) \equiv 0 \pmod{p^r}$ and set $x_r = x_{r-1} + p^{r-1}a$. Recall that $f(x_{r-1} + p^{r-1}a) \equiv f(x_{r-1}) + p^{r-1}af'(x_{r-1}) \pmod{p^r}$ by Exercise 4 of Assignment 10/05. So it suffices find an integer a such that

$$p^r | f(x_{r-1}) + p^{r-1}af'(x_{r-1}).$$

Since $p^{r-1}|f(x_{r-1})$, we have $f(x_{r-1}) = p^{r-1}b$ for some $b \in \mathbb{Z}$. The above condition then translate to

$$p|b + af'(x_{r-1}), \text{ or equivalently } [b] + [a][f'(x_{r-1})] = [0] \text{ in } \mathbb{F}_p.$$

Since $x_{r-1} \equiv x_1 \pmod{p}$, we have $f'(x_{r-1}) \equiv f'(x_1) \not\equiv 0 \pmod{p}$. So $[f'(x_{r-1})] \neq [0]$ in $\mathbb{F}_p$. But then $[f'(x_{r-1})]^{-1}$ exists in $\mathbb{F}_p$. So we can take a to be any integer whose congruence class modulo p is $-[b][f'(x_{r-1})]^{-1}$.

(uniqueness part) Suppose x_r and x_r^* both satisfy the desired condition. Then they also satisfy the same condition for $r-1$. So by the uniqueness part of $\mathcal{P}(r-1)$, we have $x_r \equiv x_r^* \pmod{p^{r-1}}$. So $x_r^* = x_r + p^{r-1}a$ for some $a \in \mathbb{Z}$. It suffices to prove $p|a$.

Now

$$0 \equiv f(x_r^*) = f(x_r + p^{r-1}a) \equiv f(x_r) + p^{r-1}af'(x_r) \equiv p^{r-1}af'(x_r) \pmod{p^r}.$$

Here the second congruence follows from Exercise 4 of Assignment 10/05. So $p^r|p^{r-1}af'(x_r)$. So $p|af'(x_r)$. So $p|a$ or $p|f'(x_r)$ since p is a prime number. But $f'(x_r) \equiv f'(x_1) \not\equiv 0 \pmod{p}$, so we must have $p|a$. Hence we are done. $\square$

9. Arithmetic in $F[x]$ (continued): Modular Arithmetic for $F[x]$

Recall that one of the major topics we covered is the comparison between $\mathbb{Z}$ and $F[x]$, where $(F, +, \cdot; 0_F, 1_F)$ is a field. They share many common properties. One thing we haven't covered yet is the modular arithmetic: we have seen this for $\mathbb{Z}$, namely the commutative ring $(\mathbb{Z}/m\mathbb{Z}, +, \cdot; [0], [1])$ for any integer $m > 1$. Recall the following properties of this ring $\mathbb{Z}/m\mathbb{Z}$.

- $(\mathbb{Z}/m\mathbb{Z}, +, \cdot; [0], [1])$ is a commutative ring.
- As a list of representatives, we can take $\mathbb{Z}/m\mathbb{Z} = \{[0], [1], ..., [m-1]\}$.
- This commutative ring is a field if and only if m is a prime number.
- $\#(\mathbb{Z}/m\mathbb{Z})^\times = \phi(m) = m\prod_{p|m}(1 - \frac{1}{p})$.

Today we do a similar construction for $F[x]$ and prove the analogues for the first three bullet points.

Let $g(x) \in F[x]$ be a nonconstant polynomial. In particular $\deg(g) \geq 1$. We say that two polynomials $f_1(x), f_2(x) \in F[x]$ are **congruent modulo** $g(x)$ and write

$$f_1(x) \equiv f_2(x) \pmod{g(x)} \quad \text{if } g(x)|f_1(x) - f_2(x).$$

In other words, $f_1(x)$ and $f_2(x)$ are congruent modulo $g(x)$ if and only if $f_1(x) - f_2(x) = g(x)h(x)$ for some $h(x) \in F[x]$. Define

$$F[x]/(g(x)) := \text{the set of congruent classes modulo } g(x).$$

For any $f(x) \in F[x]$, we denote by $\overline{f(x)}$ the congruent class of $f(x)$ modulo $g(x)$.

We define $\overline{f_1(x)} + \overline{f_2(x)} = \overline{f_1(x) + f_2(x)}$, and $\overline{f_1(x)} \cdot \overline{f_2(x)} = \overline{f_1(x)f_2(x)}$. It is not hard to check that these two binary operations are well-defined. Then it is not hard to prove:

Proposition 9.1. *$(F[x]/(g(x)), +, \cdot; \overline{0_F}, \overline{1_F})$ is a commutative ring.*

Next let us take a list of representatives for $F[x]/(g(x))$. Assume $d = \deg(g) \geq 1$ and $g(x) = \sum_{i=0}^{d} a_i x^i$ with $a_d \neq 0$. In $F[x]/(g(x))$, we have $\overline{0_F} = \overline{g(x)}$. So

$$\overline{x}^d = -a_d^{-1}(\sum_{i=0}^{d-1} a_i \overline{x}^i).$$

Hence for any polynomial $f(x) \in F[x]$, we have

$$\overline{f(x)} = \overline{f_0(x)} \text{ for some } f_0(x) \in F[x] \text{ with } \deg(f_0) \leq d-1. \tag{9.1}$$

On the other hand, if $f_1(x), f_2(x) \in F[x]$ with $\deg(f_1), \deg(f_2) < d$ satisfy $\overline{f_1(x)} = \overline{f_2(x)}$, then $f_1(x) - f_2(x) = g(x)h(x)$ for some $h(x) \in F[x]$, so

$$d > \max(\deg(f_1), \deg(f_2)) \geq \deg(f_1 - f_2) = \deg(g) + \deg(h) = d + \deg(h),$$

and hence $\deg(h) < 0$, and thus $\deg(h) = -\infty$ and $h(x) = 0$. So

$$\overline{f_1(x)} = \overline{f_2(x)} \text{ and } \deg(f_1), \deg(f_2) < d \Rightarrow f_1(x) = f_2(x). \tag{9.2}$$

By (9.1) and (9.2), we have:

Proposition 9.2. *For a list of representatives, we have*

$$F[x]/(g(x)) = \{\overline{f(x)} : \deg(f) < \deg(g)\}.$$

In particular, if F is a finite field with q elements (for notation we will denote F by $\mathbb{F}_q$)[1]*, then $\#(F[x]/(g(x))) = q^{\deg(g)}$.*

Next we prove:

Proposition 9.3. *$F[x]/(g(x))$ is a field if and only if $g(x)$ is irreducible.*

Proof. ($\Rightarrow$) Suppose $g(x)$ is not irreducible, then $g(x) = g_1(x)g_2(x)$ for some $g_1(x), g_2(x) \in F[x]$ with $0 < \deg(g_1), \deg(g_2) < \deg(g)$. But then $\overline{g_1(x)} \cdot \overline{g_2(x)} = \overline{g(x)} = \overline{0_F}$. However $\overline{g_1(x)}, \overline{g_2(x)} \neq \overline{0_F}$ by Proposition 9.2. So $F[x]/(g(x))$ is not a field if $g(x)$ is not irreducible.

($\Leftarrow$) For any $\overline{f(x)} \neq \overline{0_F}$, we have $\gcd(f(x), g(x)) = 1_F$ since $g(x)$ is irreducible. So

$$f(x)q(x) + g(x)r(x) = 1_F$$

for some $q(x), r(x) \in F[x]$. Then modulo $g(x)$ we have $\overline{f(x)} \cdot \overline{q(x)} = \overline{1_F}$. So $\overline{f(x)}$ is invertible with respect to $\cdot$ in $F[x]/(g(x))$. So every non-zero element of $F[x]/(g(x))$ is invertible, and hence $F[x]/(g(x))$ is a field. □

Comparison between $\mathbb{Z}$ and $F[x]$ (omitted in the notes).
Some exercises from Assignments (Assignment 10/05)

Exercise 1. *Let $f_1(x), f_2(x), \ldots, f_n(x) \in F[x]$ be pairwise coprime polynomials, namely there is no irreducible polynomial dividing both $f_i(x)$ and $f_j(x)$ for any $i \neq j$. Prove: if $g(x) \in F[x]$ satisfies $f_i(x)|g(x)$ for any $i = 1, \ldots, n$, then $f_1(x)f_2(x)\cdots f_n(x)|g(x)$.*

Proof. The crucial case is when $n = 2$. In general one can either imitate this case or do induction on n. Here we give a proof by induction on n.

(Base Step) $\mathcal{P}(2)$: Write $f_1(x) = c_1 p_1(x)^{\alpha_1} \cdots p_s(x)^{\alpha_s}$ and $f_2(x) = c_2 q_1(x)^{\beta_1} \cdots q_m(x)^{\beta_m}$ as the unique factorizations of $f_1(x)$ and $f_2(x)$ into monic irreducible polynomials. Since $\gcd(f_1, f_2) = 1$, we have $p_i(x) \neq q_j(x)$ for any i and j.

Write $g(x) = c r_1(x)^{\gamma_1} \cdots r_t(x)^{\gamma_t}$ as the unique factorization of $g(x)$ into monic irreducible polynomials. Since $f_1(x)|g(x)$, we know that all the $p_i(x)$'s are irreducible factors of $g(x)$. Up to reordering the $r_k(x)$'s, we may assume $r_1(x) = p_1(x)$, …, $r_s(x) = p_s(x)$. But then $\gamma_i \geq \alpha_i$ for any $i = 1, \ldots, s$ because $f_1(x)|g(x)$.

Since $f_2(x)|g(x)$ and $p_i(x) \neq q_j(x)$ for any i and j, up to reordering the $r_k(x)$'s for $k \geq s+1$, we may assume $r_{s+1}(x) = q_1(x)$, …, $r_{s+m}(x) = q_m(x)$. But then $\gamma_{s+j} \geq \beta_j$ for any $j = 1, \ldots, m$ because $f_2(x)|g(x)$.

Therefore $f_1(x)f_2(x) = c_1 c_2 \prod_{i=1}^{n} p_i(x)^{\alpha_i} \prod_{j=1}^{m} q_j(x)^{\beta_j}$ divides $\prod_{k=1}^{n+m} r_k(x)^{\gamma_k}$, which again divides $g(x)$. So $f_1(x)f_2(x)|g(x)$.

(Induction) Assume $\mathcal{P}(2), \ldots, \mathcal{P}(n-1)$ are true. We want to apply the induction hypothesis to the $n-1$ polynomials $f_1(x), \ldots, f_{n-2}(x)$ and $f_{n-1}(x)f_n(x)$. To do this it suffices to prove: for any $i = 1, \ldots, n-2$, we have $\gcd(f_i, f_{n-1}f_n) = 1$. This is true: by assumption f_i and $f_{n-1}f_n$ have no common irreducible factors for any $i = 1, \ldots, n-2$. □

Exercise 2. *Prove the following generalization of Proposition 7.12: Let F be a field. Let $f(x)$ and $g(x)$ be polynomials in $F[x]$. Suppose that $g(x)$ is irreducible and that $g'(x)$ is nonzero. Suppose r is a positive integer such that $\mathrm{char}(F) \nmid r$. Then $g(x)^{r+1}|f(x) \Leftrightarrow g(x)^r|\gcd(f(x), f'(x))$.*

[1]Next week we shall see that there exists a field with q elements if and only if $q = p^r$ for some prime number p and some integer $r \geq 1$. For each such q, there exists essentially only one field with q elements. This is why we can use $\mathbb{F}_q$ to denote **the** field with q elements.

Proof. ($\Rightarrow$) $f = g^{r+1}h$ for some $h \in F[x]$. So $f' = (r+1)g^r g'h + g^{r+1}h'$. So g^r divides both f and f', and hence divides $\gcd(f, f')$.

($\Leftarrow$) Since $g^r | f$, we have $f = g^r h$ for some $h \in F[x]$. Then $f' = rg^{r-1}g'h + g^r h'$. Since $g^r | f'$, we have $g^r | f' - g^r h' = rg^{r-1}g'h$. Since $\mathrm{char}(F) \nmid r$, we have $g^r | g^{r-1}g'h$. So $g | g'h$. Since g is irreducible, we have either $g | g'$ or $g | h$.

It suffices to prove that $g \nmid g'$. Suppose $g | g'$, then $g' = gh_0$ for some $h_0 \in F[x]$. Then $\deg(g') = \deg(g) + \deg(h_0)$. Since $g' \neq 0_F$, we have $h_0 \neq 0_F$. So $\deg(g') \geq \deg(g)$. But we have $g(x) = a_n x^n + \ldots + a_0$ with $a_n \neq 0_F$ for some $n \geq 1$, so $\deg(g) = n$ and $\deg(g') \leq n - 1$. But $\deg(g') < \deg(g)$. Hence we get a contradiction. □

Exercise 3. *Let F be a field. Let $f(x) \in F[x]$ non-constant such that $\gcd(f, f') = 1_F$. Prove that all roots of $f(x)$ in F are distinct.*

Proof. Suppose that the conclusion is not true. Then there exists some $a \in F$ such that $(x-a)^2 | f$. Then $f(x) = (x-a)^2 g(x)$ for some $g \in F[x]$. So $f' = 2(x-a)g(x) + (x-a)^2 g'(x)$. But then $x - a | f'$. So $x - a | \gcd(f, f')$. This contradicts $\gcd(f, f') = 1_F$. □

Exercise 4. *For any prime number p and positive integer r, show that*

$$f(x_0 + p^r a) \equiv f(x_0) + f'(x_0)p^r a \pmod{p^{r+1}}, \ \forall f(x) \in \mathbb{Z}[x].$$

Proof by Method 1. (Case $f(x) = a_m x^m$ for some m) In this case, $f'(x) = ma_m x^{m-1}$. So

$$f(x_0 + p^r a) = a_m(x_0 + p^r a)^m = a_m(x_0^m + mx_0^{m-1}p^r a + \sum_{i=2}^{m} \binom{m}{i} x_0^{m-i} p^{ri} a^i)$$

$$\equiv a_m(x_0^m + mx_0^{m-1}p^r a) = a_m x_0^m + ma_m x_0^{m-1} p^r a = f(x_0) + f'(x_0)p^r a \pmod{p^{r+1}}.$$

Note that this also works for $m = 0$ as $0a_0 x^{-1} = 0$ by convention (Or you can argue separately for $m = 0$ very easily).

(General case) For any $f(x) = a_n x^n + \ldots + a_0$, we have $f'(x) = na_n x^{n-1} + \ldots + a_1$. Now apply the previous case to each $a_m x^m$ for $m \in \{0, \ldots, n\}$ and do the sum, we get the desired conclusion. □

Proof by Method 2. Let $n = \deg(f)$. Since $\mathbb{Z} \subset \mathbb{R}$, we can use Taylor expansion for functions of real numbers.

$$f(x_0 + p^r a) = f(x_0) + f'(x_0)p^r a + \sum_{i=2}^{n} \frac{f^{(i)}(x_0)}{i!}(p^r a)^i$$

The right hand side is a finite sum since $f^{(i)} = 0$ for any $i > n = \deg(f)$. It suffices to prove $p^{r+1} | \frac{f^{(i)}(x_0)p^{ri}}{i!}$ for any $i \geq 2$, or equivalently to prove $p | \frac{f^{(i)}(x_0)p^{r(i-1)}}{i!}$.

This is NOT trivial as there is a denominator $i!$. Say for $i = 2$, we want to prove that $p | \frac{f''(x_0)p^r}{2}$. First of all, for this divisibility to make sense, we need to show that $\frac{f''(x_0)p^r}{2}$ is an integer. Second for $r = 1$, what we want is $p | \frac{f''(x_0)p}{2}$, or equivalently $2 | f''(x_0)$. So there is really some non-trivial work to do.

In the end the proof looks very much like Method 1: we need to use the fact that f is a polynomial. For $f(x) = a_n x^n + \ldots + a_0$, we can compute

$$f^{(i)}(x) = \sum_{m=i}^{n} a_m m(m-1)\cdots(m-i+1)x^{i-m}.$$

So

$$\frac{f^{(i)}(x_0)}{i!} = \sum_{m=i}^{n} a_m \binom{m}{i} x_0^{i-m}$$

is an integer. Hence $p|\frac{f^{(i)}(x_0)}{i!}p^{r(i-1)}$ for any $i \geq 2$. □

10. Classification of Finite Fields

10.1. **Some preparation on abelian groups.** Let $(G, \cdot; e)$ be an abelian group.

Definition 10.1. *For any element $a \in G$, the order of a, denoted by* $\mathrm{ord}(a)$*, is defined to be the smallest positive integer d such that $a^d = e$. If such a d does not exist we say that a has order $+\infty$.*

Lemma 10.2. *Suppose $m \neq 0$ is an integer. Then the order of a^m is* $\mathrm{ord}(a)/\gcd(m, \mathrm{ord}(a))$ *if* $\mathrm{ord}(a) < +\infty$ *and is* $+\infty$ *if* $\mathrm{ord}(a) = +\infty$.

Lemma 10.3. *Let G be a finite group such that $\#G = n$. Then the order of any element $a \in G$ divides n.*

These two lemmas will be your homework today.

As an application, we prove the following proposition on numbers.

Proposition 10.4. *Let n be a positive integer. Then $n = \sum_{d|n,\ d>0} \phi(d)$.*

Proof. Consider the abelian group $(\mathbb{Z}/n\mathbb{Z}, +; [0])$.

For any positive divisor d of n, the equation $dx = [0]$ has exactly d solutions, namely $[\frac{n}{d}y]$ for $y = 0, \ldots, d-1$. These are all the elements of $\mathbb{Z}/n\mathbb{Z}$ of order dividing d, and Lemma 10.2 implies that there are exactly $\phi(d)$ among them of order d (namely those y such that $\gcd(y, d) = 1$).

By Lemma 10.3, every element of $\mathbb{Z}/n\mathbb{Z}$ has order dividing n. So the number of elements of $\mathbb{Z}/n\mathbb{Z}$ is the sum of the number of elements of order d for each $d|n$, $d > 0$. In other words $\#(\mathbb{Z}/n\mathbb{Z}) = \sum_{d|n,\ d>0} \#\{a \in \mathbb{Z}/n\mathbb{Z} : \mathrm{ord}(a) = d\}$.

Combine the last two paragraphs and we can conclude. □

10.2. **Basic properties of a finite field F.** Let $(F, +, \cdot; 0_F, 1_F)$ be a field. Recall the definition of the characteristic of F: we say that $\mathrm{char}(F) = n$ if n is the smallest positive integer such that $n1_F = 0_F$. If such an integer does not exist, we say that $\mathrm{char}(F) = 0$.

Denote by $\varphi\colon \mathbb{Z} \to F$, $n \mapsto n \cdot 1_F$.

Recall that for any field F, either $\mathrm{char}(F) = 0$ or $\mathrm{char}(F) = p$ is a prime number. If $\mathrm{char}(F) = 0$, then F has infinitely many elements because the map φ is injective (and hence induces an injective map $\mathbb{Q} \to F$, $m/n \mapsto (m \cdot 1_F)(n \cdot 1_F)^{-1}$). If $\mathrm{char}(F) = p$, then φ induces an injection $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z} \to F$, $[n] \mapsto n \cdot 1_F$. In other words, $\mathbb{F}_p$ is the smallest field of characteristic p, and every field with p elements is isomorphic to $\mathbb{F}_p$.

Theorem 10.5. *Suppose F has q elements. Then q is a power of p where the prime number $p = \mathrm{char}(F)$.*

Proof. Since F is a finite field, its characteristic is a prime number p. So $p \cdot a = 0_F$ for any $a \in F$.

We will construct an ascending chain of subsets of F which are closed under addition and negation

$$V_1 \subset V_2 \subset \ldots \subset V_i \subset \ldots$$

such that $\#V_i = p^i$. Since F has only finitely many elements, there exists some $r \geq 1$ such that $\#F < p^{r+1}$. But then we must have $F = V_r$ and hence $\#F = p^r$ (and the chain stops at V_r). We are done.

Let us construct the chain by induction.

Let a_1 be a non-zero element of F. Define $V_1 := \{n_1 \cdot a_1 : n_1 \in \mathbb{Z}\}$. Then V_1 is clearly closed under addition and negation. We prove that $\#V_1 = p$: For any two numbers $m_1, m_1' \in \{0, \dots, p-1\}$, if $m_1 \cdot a_1 = m_1' \cdot a_1$, then $(m_1 - m_1')a_1 = 0_F$ and hence $((m_1 - m_1')1_F)a_1 = 0_F$. Since $a_1 \neq 0_F$, we have $(m_1 - m_1')1_F = 0_F$. So $p | m_1 - m_1'$. But $-(p-1) \leq m_1 - m_1' \leq p-1$, so $m_1 = m_1'$. Thus $\#V_1 = p$.

Assume $V_1, \dots, V_{i-1}$ are constructed ($i \geq 2$). If $F = V_{i-1}$, then we are done. Otherwise let $a_i \in F \setminus V_{i-1}$. Let $V_i = \{v + n_i a_i : v \in V_{i-1},\ n_i \in \mathbb{Z}\}$. Then V_i is clearly closed under addition and negation.We prove that $\#V_i = p^i$: For any $v, v' \in V_{i-1}$ and any numbers $m_i, m_i' \in \{0, \dots, p-1\}$, if $v + m_i a_i = v' + m_i' a_i$, then $(m_i - m_i')a_i = v' - v \in V_{i-1}$. Since $-(p-1) \leq m_i - m_i' \leq p-1$, we have either $m_i = m_i'$ or $\gcd(m_i - m_i', p) = 1$. In the latter case, we have $s(m_i - m_i') + tp = 1$ for some $s, t \in \mathbb{Z}$. But then $a_i = s(m_i - m_i')a_i + tpa_i = s(m_i - m_i')a_i \in V_{i_1}$, contradicts $a_i \notin V_{i-1}$. Thus $m_i = m_i'$ and hence $v = v'$. Therefore $\#V_i = (\#V_{i-1})p = p^i$. □

Theorem 10.6. *Suppose F has q elements. Then $(F^\times, \cdot; 1_F)$ is isomorphic to $(\mathbb{Z}/(q-1)\mathbb{Z}, +; [0])$ as abelian groups, namely there exists a bijective map $\psi\colon F^\times \to \mathbb{Z}/(q-1)\mathbb{Z}$ such that $\psi(1_F) = [0]$ and $\psi(ab) = \psi(a) + \psi(b)$ for any $a, b \in F^\times$.*

Proof. For simplicity we write n for $q-1$. Then $F^\times$ is an abelian group with n elements. So by Ex 2 of Assignment 09/21, we have $a^n = 1_F$ for any $a \in F^\times$. In other words, every element of $F^\times$ is a root of the polynomial $x^n - 1_F \in F[x]$.

For any element $a \in F^\times$, define a map

$$\psi_a\colon \mathbb{Z}/n\mathbb{Z} \to F^\times,\ [m] \mapsto a^m.$$

Check that this map is well-defined and is a group homomorphism (namely $\psi_a([0]) = 1_F$ and $\psi_a([m] + [m']) = \psi_a([m])\psi([m'])$ for any $[m], [m'] \in \mathbb{Z}/n\mathbb{Z}$). Then it suffices to prove that ψ_a is bijective for some $a \in F^\times$, or equivalently to prove that $\mathrm{ord}(a) = n$ for some $a \in F^\times$.

Since $F^\times$ is an abelian group with n elements, we have $\mathrm{ord}(a) | n$ for any $a \in F^\times$ by Lemma 10.3.

Suppose $d | n$ and $d > 0$. We will prove

$$F^\times \text{ contains at most } \phi(d) \text{ elements of order } d. \tag{10.1}$$

The polynomial $x^d - 1_F$ has degree d and hence at most d roots in F. In other words, there are at most d elements whose order divides d.

Now if a satisfies $\mathrm{ord}(a) = d$, then $1_F = a^0, a, a^2, \dots, a^{d-1}$ all satisfy $x^d - 1_F = 0_F$. Hence we have found d elements of $F^\times$ whose order divides d. Thus the set of elements of $F^\times$ whose order divides d is exactly $\{1_F, a, a^2, \dots, a^{d-1}\}$.[2]

Any element in the set $\{1_F, a, a^2, \dots, a^{d-1}\}$ whose order is exactly d is of the form a^m for some $m \in \{0, 1, \dots, d-1\}$ such that $\gcd(m, d) = 1$. This follows from Lemma 10.2. There are $\phi(d)$ such integers. Hence we have proven (10.1).

Thus the total number of elements in $F^\times$ whose order is smaller than n is at most $\sum_{d|n, 1 \leq d < n} \phi(d)$. So the number of elements in $F^\times$ of order n is $n - \sum_{d|n, 1 \leq d < n} \phi(d)$, which equals $\phi(n) > 0$ by Proposition 10.4. Hence we are done. □

[2] *Attention* This is different from the proof of Proposition 10.4: in the current proof, such an element a may NOT exist a priori, so that $F^\times$ may NOT have any element of order dividing d. It is part of the proof to establish the existence of such an a.

11. Classification of Finite Fields (Continued)

11.1. **Algebraically closed field.**

Definition 11.1. *A field $(F, +, \cdot; 0_F, 1_F)$ is said to be **algebraically closed** if every non-constant polynomial in $F[x]$ has at least one root in F.*

The field $\mathbb{Q}$ is not algebraically closed since the polynomial $x^2 - 2$ has no roots in $\mathbb{Q}$. The field $\mathbb{R}$ is not algebraically closed since the polynomial $x^2 + 1$ has no roots in $\mathbb{R}$. The Fundamental Theorem of Algebra states that $\mathbb{C}$ is algebraically closed.

On the other hand, $\mathbb{F}_3$ is not algebraically closed since the polynomial $x^2 - 2$ has no roots in $\mathbb{F}_3$. In fact, no finite field is algebraically closed: suppose F is a finite field with q elements, then $(F^\times, \cdot; 1_F)$ is an abelian group with $q-1$ elements, so $x^q - x = 0_F$ for any $x \in F$, and hence the polynomial $x^q - x + 1_F$ has no roots in F.

Theorem 11.2. *For any field F, there exists an algebraically closed field $\overline{F}$ which contains F as a subfield.*

Such an $\overline{F}$ is not unique. The proof of this theorem uses foundational assumption in logic theory. The basic idea is to add all roots of all polynomials in $F[x]$ and so on. We will not present it here. But we make the following observation: $0_F \in F \subset \overline{F}$ is the unit for addition on $\overline{F}$ and $1_F \in F \subset \overline{F}$ is the unit for multiplication on $\overline{F}$. So we can use $(\overline{F}, +, \cdot; 0_F, 1_F)$ to denote this algebraically closed field (no need to introduce the notation $0_{\overline{F}}$ and $1_{\overline{F}}$).

Theorem 11.3. *Let F be an algebraically closed field. Let $f \in F[x]$ be a polynomial of degree $d \geq 0$. Then f has exactly d roots in F counted with multiplicity. Namely*

(1) $f(x) = c(x - x_1) \dots (x - x_d)$ for some elements $c, x_1, \dots, x_d \in F$;
(2) if $f(x) = c(x - x_1) \dots (x - x_d) = c'(x - x'_1) \dots (x - x'_d)$, then $c = c'$ and $\{x_1, \dots, x_d\} = \{x'_1, \dots, x'_d\}$ as multi-sets.

Proof. (1) Induction of d.

(base step) $\mathcal{P}(0)$ and $\mathcal{P}(1)$ are clearly true.

(induction step) Assume $\mathcal{P}(1), \dots, \mathcal{P}(d-1)$ are true. By definition of algebraically closed field, f has a root $x_d \in F$. So $x - x_d | f$. So $f = (x - x_d)g$ for some $g \in F[x]$. So $d = \deg(f) = 1 + \deg(g)$. But then $\mathcal{P}(d-1) \Rightarrow g(x) = c(x - x_1) \dots (x - x_{d-1})$ for some $c, x_1, \dots, x_{d-1} \in F$. Hence we are done.

(2) This follows directly from UFT since any degree 1 polynomial is irreducible.

□

This theorem completes Proposition 6.8. Moreover in view of Theorem 11.2, it implies Proposition 6.8 immediately.

Before going on, let us state the following result which is an immediate corollary of Theorem 11.3 (or even of Proposition 6.8).

Corollary 11.4. *Let F be a field. If $f, g \in F[x]$ of the same degree d satisfy that $f(a) = g(a)$ for at least $d + 1$ elements $a \in F$, then $f = g$.*

This tells us when we can decide whether two polynomials are equal by testing the evaluations. In particular if F is infinite (for example when $\mathrm{char}(F) = 0$ or F is algebraically closed), then two polynomials are equal if they have the same evaluations on every element of F.

11.2. **The field $\mathbb{F}_q$.** Let p be a prime number. We fix an algebraic closed field $\overline{\mathbb{F}}_p$ which contains $\mathbb{F}_p$ as a subfield. For simplicity, we will omit the "[]" for [0] and [1]. So we have a field inclusion $(\mathbb{F}_p, +, \cdot; 0, 1) \subset (\overline{\mathbb{F}}_p, +, \cdot; 0, 1)$.

Last time we have seen that the cardinality of any finite field of characteristic p is p^r for some integer $r \geq 1$. Conversely, let $q = p^r$. The polynomial $g(x) = x^q - x$ then has q roots in $\overline{\mathbb{F}}_p$ by Theorem 11.3. Since $g'(x) = qx^{q-1} - 1 = -1$ is coprime to $g(x)$, we see that all the q roots of $g(x)$ are distinct by Proposition 7.12 (i.e. no roots with multiplicity at least 2). Let $\mathbb{F}_q$ denote the set of all the roots of $g(x)$. This is a set with q elements. Then $\mathbb{F}_p \subset \mathbb{F}_q$ as sets by Fermat's Little Theorem.

It is easy to check that $\mathbb{F}_q$ is closed under negation, multiplication and inverse. Checking that $\mathbb{F}_q$ is closed under addition needs some extra work. We need to use Ex 3 of Assignment 10/17 ($(a+b)^q = a^q + b^q$ for any $a, b \in \overline{\mathbb{F}}_p$). Hence we obtain a field $(\mathbb{F}_q, +, \cdot; 0, 1)$ and the following field inclusions

$$(\mathbb{F}_p, +, \cdot; 0, 1) \subset (\mathbb{F}_q, +, \cdot; 0, 1) \subset (\overline{\mathbb{F}}_p, +, \cdot; 0, 1).$$

This proves the following theorem.

Theorem 11.5. *For any $q = p^r$ for some prime number p and some integer $r \geq 1$, there exists a field with q elements.*

$\boxed{Useful\ Fact}$ For any $x \in \overline{\mathbb{F}}_p$, we have $x \in \mathbb{F}_q \Leftrightarrow x^q = x$.

In fact, we have a much stronger uniqueness result, which is a particular property of finite fields.

Theorem 11.6. *Let $q = p^r$ for some prime number p and some integer $r \geq 1$. Then every field F with q elements is isomorphic to $\mathbb{F}_q$.*

Proof. Recall that by Theorem 10.6, as an abelian group we have an isomorphism $\varphi\colon (F^\times, \cdot; 1_F) \xrightarrow{\sim} (\mathbb{Z}/(q-1)\mathbb{Z}, +; [0])$. Let $a = \varphi^{-1}([1])$, then $\mathrm{ord}(a) = q - 1$. So $F^\times = \{1, a, ..., a^{q-1}\}$.

For any positive integer $1 \leq i \leq q-1$, let $V_i = \{\sum_{j=0}^{i-1} n_j a^j : 0 \leq n_j \leq p-1\}$. Then $V_1 \subset V_2 \subset ... \subset V_{q-1}$, and $F = V_{q-1}$. Moreover,

- $\#V_i \leq p^i$, with "=" if and only if $a^{i-1} \notin V_{i-1}$.
- If $a^{i-1} \in V_{i-1}$, then $V_{i-1} = V_i = ... = V_{q-1} = F$.

But $\#V_{r+1} \leq \#F = p^r < p^{r+1}$, so $a_r \in V_r$. In other words, there exist $c_0, ..., c_{r-1} \in \mathbb{F}_p$ such that

$$c_0 + ... + c_{r-1}a^{r-1} + a^r = 0.$$

Let $g(x) = c_0 + ... + c_{r-1}x^{r-1} + x^r \in \mathbb{F}_p[x]$. Then $\deg(g) = r$.

We prove that $F = V_r$. It suffices to prove $\#V_r = p^r$. Suppose not, then $a^{r-1} \in V_{r-1}$, and so $V_{r-1} = V_r = ... = V_{q-1} = F$. But $\#V_{r-1} \leq p^{r-1}$ and $\#F = p^r$, contradiction!

Now we are ready to define the map (identify $F = V_r$)

$$\psi\colon F \to \mathbb{F}_p[x]/(g(x)), \quad \sum_{i=0}^{r-1} n_i a^i \mapsto \overline{\sum_{i=0}^{r-1} n_i x^i}.$$

It is not hard to check that this map is a ring homomorphism and is surjective. Hence ψ is also injective because $\#F = \#(\mathbb{F}_p[x]/(g(x))) = p^r$. So ψ is a ring isomorphism. But F is a field, so $\mathbb{F}_p[x]/(g(x))$ is also a field. So $g(x)$ is irreducible.

To sum it up, we have a field isomorphism $\psi\colon F \to \mathbb{F}_p[x]/(g(x))$ where $g(x) \in \mathbb{F}_p[x]$ is an irreducible polynomial.

Next we prove that there exists an injective ring homomorphism

$$\iota\colon \mathbb{F}_p[x]/(g(x)) \to \overline{\mathbb{F}}_p.$$

Let α be a root of $g(x)$ in $\overline{\mathbb{F}}_p$ and define ι to be the map sending each polynomial $\overline{h(x)}$ to $h(\alpha)$. Then ι is well-defined as $g(\alpha) = 0$. This map ι is injective: let $h(x)$ a monic polynomial in $\mathbb{F}_p[x]$ of minimal degree such that $h(\alpha) = 0$, then $g(x) = h(x)q(x) + r(x)$ where $\deg(r) < \deg(h)$, and so $0 = g(\alpha) = h(\alpha)q(\alpha) + r(\alpha) = r(\alpha)$. But then $r(x)$ is a polynomial such that $r(\alpha) = 0$ and $\deg(r) < \deg(h)$, contradicting the minimality of $\deg(h)$ unless $r(x) = 0$. So $r(x) = 0$ and so $h(x)|g(x)$. But $g(x)$ is irreducible, so either $h(x) = 1$ or $h(x) = g(x)$. But $h(\alpha) = 0$, so $h(x) = g(x)$ and hence $\overline{h(x)} = \overline{0}$.

Therefore we have proven that there exists an injective field homomorphism $\iota \circ \psi\colon F \to \overline{\mathbb{F}}_p$. It suffices to prove that $\iota \circ \psi(F) = \mathbb{F}_q$. But $\iota \circ \psi(F)$ is a subfield of $\overline{\mathbb{F}}_p$ having q elements, so it suffices to prove that $\mathbb{F}_q$ is the unique subfield of $\overline{\mathbb{F}}_p$ with q elements. This is true: every element of a field with q elements satisfies $x^q = x$ and $\mathbb{F}_q$ is the set of all the roots of $x^q - x$. □

11.3. **Summary.** Let us make a summary of the results concerning the classification of finite fields with we have obtained so far. See Theorem 10.5, Theorem 11.5 and Theorem 11.6.

(1) Any finite field has characteristic p for a prime number p.
(2) The cardinality of any finite field F is p^r for $p = \mathrm{char}(F)$ and some integer $r \geq 1$.
(3) (existence) For any $q = p^r$ for some prime number p and some integer $r \geq 1$, there exists a finite field F with q elements.
(4) (uniqueness) Moreover any two finite fields with q elements are isomorphic to each other.

12. Solutions to Assignment 10/10 and 10/12

Exercise 1. *For any non-constant polynomial $f(x) \in \mathbb{Q}[x]$ with $f(x) = cp_1(x)^{\alpha_1} \cdots p_n(x)^{\alpha_n}$ as the factorization into monic irreducible polynomials as above, prove* $\deg(f) - \deg(\mathrm{rad}(f)) \leq \deg(\gcd(f(x), f'(x)))$. *(Hint: prove $p_1(x)^{\alpha_1 - 1} \cdots p_n(x)^{\alpha_n - 1} | \gcd(f(x), f'(x))$.)*

Proof. (Because $p_i(x)$ are irreducible polynomials, they are non-constant by definition. So $p_i'(x)$ is nonzero for any $i = 1, \ldots, n$ because $p_i(x) \in \mathbb{Q}[x]$ ($\mathbb{Q}$ is of characteristic 0!).)[3] By the "only if" part of Proposition 7.12 and its generalization (i.e. Exercise 2 of Assignment 10/05)[4], we have $p_i(x)^{\alpha_i - 1} | \gcd(f(x), f'(x))$. But the $p_i(x)$'s are pairwise coprime because they are all irreducible. So By Exercise 1 of Assignment 10/05, we have $p_1(x)^{\alpha_1 - 1} \cdots p_n(x)^{\alpha_n - 1} | \gcd(f(x), f'(x))$. But then

$$\begin{aligned}
\deg(f) - \deg(\mathrm{rad}(f)) &= \sum_{i=1}^{n} \alpha_i \deg(p_i) - \sum_{i=1}^{n} \deg(p_i) \\
&= \sum_{i=1}^{n} (\alpha_i - 1) \deg(p_i) \\
&= \sum_{i=1}^{n} \deg(p_i(x)^{\alpha_i - 1}) \\
&= \deg(p_1(x)^{\alpha_1 - 1} \cdots p_n(x)^{\alpha_n - 1}) \\
&\leq \deg(\gcd(f(x), f'(x))) \quad \text{since } \gcd(f(x), f'(x)) \neq 0.
\end{aligned}$$

Another proof: if we compute $f'(x)$ directly, then we see that $\prod_{i=1}^{n} p_i(x)^{\alpha_i - 1}$ divides every term in the sum, so $\prod_{i=1}^{n} p_i(x)^{\alpha_i - 1} | f'(x)$. So $\prod_{i=1}^{n} p_i(x)^{\alpha_i - 1} | \gcd(f, f')$. But $f(x)$ is non-constant, so $\gcd(f, f') \neq 0$. So

$$\deg(f) - \deg(\mathrm{rad}(f)) = \deg\left(\frac{f(x)}{\mathrm{rad}(f)}\right) = c \prod_{i=1}^{n} p_i(x)^{\alpha_i - 1} \leq \deg(\gcd(f, f')).$$

□

Exercise 2. *(Mason's Theorem). Let $f(x)$, $g(x)$ and $h(x)$ be non-constant polynomials in $\mathbb{Q}[x]$ such that no irreducible polynomial divides all three of them. Suppose $f(x) + g(x) = h(x)$. We prove the following inequality:*

$$\max\{\deg(f), \deg(g), \deg(h)\} \leq \deg(\mathrm{rad}(f(x)g(x)h(x))) - 1.$$

1. *Prove that f, g and h are pairwise coprime. Deduce from this that* $\gcd(f, f')$, $\gcd(g, g')$ *and* $\gcd(h, h')$ *are pairwise coprime.*
2. *Prove: $f'(x)g(x) - g'(x)f(x) = f'(x)h(x) - h'(x)f(x)$. Deduce from this that* $\gcd(f, f')$, $\gcd(g, g')$ *and* $\gcd(h, h')$ *all divide $f'(x)g(x) - g'(x)f(x)$.*
3. *Prove:* $\gcd(f, f')\gcd(g, g')\gcd(h, h') | f'(x)g(x) - g'(x)f(x)$.

---

[3] This is not necessary by the next footnote.

[4] The statement is the following: Let F be a field. Let $f(x)$ and $g(x)$ be polynomials in $F[x]$. Suppose that $g(x)$ is irreducible and that $g'(x)$ is nonzero. Suppose r is a positive integer such that $\mathrm{char}(F) \nmid r$. Then $g(x)^{r+1}$ divides $f(x)$ if and only if $g(x)^r$ divides $\gcd(f(x), f'(x))$. But the "only if" part does NOT need $g'(x) \neq 0_F$ or $\mathrm{char}(F) \nmid r$.

(4) *Prove the following inequalities:*

$$\begin{aligned}&\deg(f)+\deg(g)+\deg(h)-\deg(\mathrm{rad}(f))-\deg(\mathrm{rad}(g))-\deg(\mathrm{rad}(h))\\ \leq&\deg(f'(x)g(x)-g'(x)f(x))\\ \leq&\deg(f)+\deg(g)-1.\end{aligned}$$

(5) *Prove:*

$$\begin{aligned}&\deg(h)\\ \leq&\deg(\mathrm{rad}(f))+\deg(\mathrm{rad}(g))+\deg(\mathrm{rad}(h))-1\\ =&\deg(\mathrm{rad}(f)\mathrm{rad}(g)\mathrm{rad}(h))-1\\ =&\deg(\mathrm{rad}(fgh))-1.\end{aligned}$$

(6) *Conclude by rewriting the equation $f+g=h$ as $f+(-h)=-g$ (so that "$g(x)$" becomes "$-h(x)$" and "$h(x)$" becomes "$-g(x)$") and $g+(-h)=-f$. Your argument for this question should not exceed four lines.*

Proof. (1) Suppose not. Say $r(x)$ is an irreducible polynomial dividing two of $f(x), g(x), h(x)$. Then since $f(x)+g(x)=h(x)$, we have that $r(x)$ divides the third one. Hence $r(x)$ divides all three of them. This is a contradiction to our hypothesis.

Now $f(x)$, $g(x)$, $h(x)$ are pairwise coprime, so they have different irreducible factors. So do $\gcd(f,f')$, $\gcd(g,g')$ and $\gcd(h,h')$. Hence $\gcd(f,f')$, $\gcd(g,g')$ and $\gcd(h,h')$ are also pairwise coprime.

(2) $f'h-fh'=f'(f+g)-f(f+g)'=f'(f+g)-f(f'+g')=f'g-fg'$.

It is clear that $\gcd(f,f')|f'g-fg'$ and $\gcd(g,g')|f'g-fg'$. On the other hand, $\gcd(h,h')|f'h-fh'=f'g-fg'$.

(3) This follows directly from (1), (2) and Exercise 1 of Assignment 10/05.

(4) Apply Exercise 1 to $f(x)$, $g(x)$ and $h(x)$ respectively, we have

$$\begin{aligned}&\deg(f)-\deg(\mathrm{rad}(f))+\deg(g)-\deg(\mathrm{rad}(g))+\deg(h)-\deg(\mathrm{rad}(h))\\ \leq&\deg(\gcd(f,f'))+\deg(\gcd(g,g'))+\deg(\gcd(h,h'))\\ =&\deg(\gcd(f,f')\gcd(g,g')\gcd(h,h'))\end{aligned}$$

If $f'g-fg'\neq 0$, then by (3) we have $\deg(\gcd(f,f')\gcd(g,g')\gcd(h,h'))\leq\deg(f'g-fg')$. Now by definition of formal derivative, we have $\deg(f')\leq\deg(f)-1$ and $\deg(g')\leq\deg(g)-1$ since f and g are non-constant. So $\deg(f'g-fg')\leq\max\{\deg(f'g),\deg(fg')\}\leq\deg(f)+\deg(g)-1$.

Thus we are left to prove $f'g-fg'\neq 0$. Let $p(x)^\alpha$ be an irreducible factor of $f(x)$ with maximal power, meaning $p(x)^\alpha|f(x)$ but $p(x)^{\alpha+1}\nmid f(x)$. So $f(x)=p(x)^\alpha f_0(x)$ for some $f_0(x)\in\mathbb{Q}[x]$ with $p(x)\nmid f_0(x)$. So $f'(x)=\alpha p(x)^{\alpha-1}f_0(x)+p(x)^\alpha f_0'(x)$. Now $p(x)^\alpha\nmid\alpha p(x)^{\alpha-1}f_0(x)$ since $p(x)\nmid f_0(x)$ and $\alpha p(x)^{\alpha-1}f_0(x)\neq 0$ ($f(x)$ is non-constant and characteristic 0!), so $p(x)^\alpha\nmid f'(x)$. So we have $p(x)^\alpha|fg'$ and $p(x)^\alpha\nmid f'g$ since f and g are coprime. So $f'g-fg'\neq 0$.

(5) By (4), we have

$$\deg(h)\leq\deg(\mathrm{rad}(f))+\deg(\mathrm{rad}(g))+\deg(\mathrm{rad}(h))-1=\deg(\mathrm{rad}(f)\mathrm{rad}(g)\mathrm{rad}(h))-1.$$

But f, g and h are pairwise coprime, so they have different irreducible factors. So by definition of radical, we have $\mathrm{rad}(f)\mathrm{rad}(g)\mathrm{rad}(h) = \mathrm{rad}(fgh)$. Hence we are done.

(6) By assumption, $f+(-h) = -g$, and f, $-h$, $-g$ are non-constant such that no irreducible polynomial divides all three of them. Apply the conclusion of (5) to f, $-h$ and $-g$, we have $\deg(g) = \deg(-g) \leq \deg(\mathrm{rad}(f)\mathrm{rad}(g)\mathrm{rad}(h)) - 1$. Similarly from $g + (-h) = -f$ we obtain $\deg(f) = \deg(-f) \leq \deg(\mathrm{rad}(f)\mathrm{rad}(g)\mathrm{rad}(h)) - 1$. Hence we are done.

□

Exercise 3. *Formulate the analogous statement of Mason's Theorem, namely Exercise 3, for $\mathbb{Z}$ (i.e. replace $\mathbb{Q}[x]$ by $\mathbb{Z}$). This statement is called the abc conjecture. (Hint: start by defining the radical for any integer). You only need to write down this statement and no explanation or proof is required. After you finish all the other exercises including those below, you can try to prove the abc conjecture.*

Proof. Let a, b and c be positive integers such that $a + b = c$ and no prime number divides all three of them. Then $c < \mathrm{rad}(abc)$.[5] Here $\mathrm{rad}(\prod_{i=1}^{r} p_i^{\alpha_i}) = \prod_{i=1}^{r} p_i$ for any prime numbers p_i and integers $\alpha_i \geq 1$. □

Exercise 4. *The goal of this exercise is to prove the following result (known as Fermat's Last Theorem for $\mathbb{Q}[x]$): Let $n \geq 3$ be an integer. Then there are no non-constant polynomials $f(x)$, $g(x)$ and $h(x)$ in $\mathbb{Q}[x]$ such that there is no irreducible polynomial dividing all three of them and they satisfy*

$$f(x)^n + g(x)^n = h(x)^n.$$

(Hint: apply Mason's Theorem, namely Exercise 2, to $f(x)^n$, $g(x)^n$ and $h(x)^n$).

Proof. Suppose such $f(x)$, $g(x)$, $h(x)$ exist. Apply Mason's Theorem to $f(x)^n$, $g(x)^n$ and $h(x)^n$, we have

$$n \max\{\deg(f), \deg(g), \deg(h)\} \leq \deg(\mathrm{rad}(f^n g^n h^n)) - 1 = \deg(\mathrm{rad}(fgh)) - 1.$$

But $\deg(\mathrm{rad}(fgh)) \leq \deg(fgh) = \deg(f) + \deg(g) + \deg(h)$, so

$$\begin{aligned} &3 \max\{\deg(f), \deg(g), \deg(h)\} \\ \leq &n \max\{\deg(f), \deg(g), \deg(h)\} \\ \leq &\deg(f) + \deg(g) + \deg(h) - 1 \\ \leq &3 \max\{\deg(f), \deg(g), \deg(h)\} - 1. \end{aligned}$$

But this cannot happen. So we get a contradiction and we are done. □

Exercise 5. *Formulate the analogue of Fermet's Last Theorem for $\mathbb{Q}[x]$, namely Exercise 5, for $\mathbb{Z}$ (i.e. replace $\mathbb{Q}[x]$ by $\mathbb{Z}$). This statement is called Fermat's Last Theorem. You only need to write down this statement and no explanation or proof is required.*

A bonus will be given if you use the abc conjecture, namely the statement that you formulated in Exercise 3, to prove Fermat's Last Theorem for $\mathbb{Z}$.

[5] In fact we need to include a calibration factor for this inequality. But for the purpose of this course (compare $\mathbb{Z}$ and $F[x]$), this is good enough. The correct conclusion of the *abc* conjecture should be: For any $\epsilon > 0$, there exists a real number $K_\epsilon > 0$ such that $c < K_\epsilon \mathrm{rad}(abc)^{1+\epsilon}$.

Proof. For any integer $n \geq 3$, there are no positive integers a, b, c such that no prime number divides all three of them and

$$a^n + b^n = c^n.$$

Apply the *abc* conjecture to a^n, b^n and c^n (the formulation in Exercise 3):

$$c^n < \mathrm{rad}(a^n b^n c^n) = \mathrm{rad}(abc) \leq abc \leq c^3.$$

This contradicts $n \geq 3$. □

Exercise 6. *Briefly explain why our statements for Mason's Theorem and Fermat's Last Theorem for $\mathbb{Q}[x]$ are wrong when $\mathbb{Q}[x]$ is replaced by $\mathbb{F}_p[x]$. You can do this either by pointing out which step of your proof does not work for $\mathbb{F}_p[x]$ or by giving a counterexample. A bonus will be given if you can furthermore correct the statements for $\mathbb{F}_p[x]$ (no need to prove this correction).*

Proof. This is because of the last paragraph of the proof of Step (4): the fact that $f(x) \in \mathbb{F}_p[x]$ is non-constant does NOT imply $f'(x) \neq 0$. So it may happen that $f'g - fg' = 0$ (in which case both f' and g' must be zero).

To correct the statement, simply replace "$f(x)$, $g(x)$, $h(x)$ are non-constant" by "$f'(x) \neq 0$, $g'(x) \neq 0$ and $h'(x) \neq 0$". □

13. Quadratic Reciprocity

We start our topic on solving Diophantine equations now. The first non-trivial question we ask is: what integers can be represented as a sum of two squares? In other words, we are looking for the integer solutions to $z = x^2 + y^2$.

From the Unique Factorization Theorem, it is helpful to know what prime factors can a sum of two square have. That is, we want to ask when the equation

$$x^2 + y^2 \equiv 0 \pmod{p}$$

has a solution.

There is an obvious solution, namely $x, y \equiv 0 \pmod{p}$. We call this the trivial solution. Note that in this case, $x^2 + y^2$ will be divisible not just by p, but also by p^2. Suppose now we have a non-trivial solution. Then neither x nor y can be congruent to 0 modulo p. Dividing by x^2 then implies that -1 is a square in $\mathbb{F}_p$.

13.1. Quadratic residue v.s. nonresidue.

Definition 13.1. *Let m be a positive integer and let a be an integer. Then a is **quadratic residue modulo** m if the equation $x^2 \equiv a \pmod{m}$ has a solution. Otherwise, a is called a **quadratic nonresidue**.*

Our question now becomes: when is -1 a quadratic residue modulo p? Since every element of $\mathbb{F}_2$ is a square, we may assume $p > 2$ is odd. Suppose -1 is a quadratic residue modulo p, then there exists $a \in \mathbb{F}_p$ such that $a^2 = -1$ in $\mathbb{F}_p$. Clearly, a is nonzero and so Fermat's Little Theorem implies $a^{p-1} \equiv 1 \pmod{p}$. On the other hand, since p is odd, $p - 1$ is divisible by 2 and we have

$$a^{p-1} = (a^2)^{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2}} \pmod{p}.$$

Hence we conclude that $\frac{p-1}{2}$ is even and hence $p \equiv 1 \pmod{4}$.

Conversely, suppose $p \equiv 1 \pmod{4}$. Let $\alpha \in \mathbb{F}_p^\times$ be a primitive $(p-1)$-th root of unity (it exists by Theorem 10.6). Then there exists an integer m such that $\alpha^m = -1$. Since $(-1)^{\frac{p-1}{2}} = 1$, we see that $\alpha^{\frac{m(p-1)}{2}} = 1$ and hence $\frac{m(p-1)}{2}$ is divisible by $p - 1$. This implies that m is even. Taking $a = \alpha^{\frac{m}{2}}$ then gives a solution to $x^2 = -1$ in $\mathbb{F}_p$.

This proves the following criterion.

Proposition 13.2. *-1 is a square modulo p if and only if $p = 2$ or $p \equiv 1 \pmod{4}$.*

In fact the above argument generalizes to give a proof of the following result:

Proposition 13.3. *Let $p > 2$ be a prime number. Then an integer a is a quadratic residue modulo p if and only if $a^{\frac{p-1}{2}} = 1$ in $\mathbb{F}_p$. Equivalently, an integer a is a quadratic nonresidue modulo p if and only if $a^{\frac{p-1}{2}} = -1$ in $\mathbb{F}_p$.*

Definition 13.4. *Let p be a prime number and let a be an integer. The **Legendre symbol** is defined to be*

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } p \nmid a \text{ and } a \text{ is a quadratic residue modulo } p \\ 0 & \text{if } p | a \\ -1 & \text{if } a \text{ is not a quadratic residue modulo } p \end{cases}.$$

In other words,

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } [a] \neq [0] \text{ is a square in } \mathbb{F}_p \\ 0 & \text{if } [a] = [0] \text{ in } \mathbb{F}_p \\ -1 & \text{if } [a] \text{ is not a square in } \mathbb{F}_p \end{cases}.$$

Note that by Proposition 13.3, we have

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

for any prime number $p > 2$. This shows that the Legendre symbol is multiplicative, namely

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \tag{13.1}$$

Now suppose we want to compute $\left(\frac{n}{p}\right)$ for an integer $n \in \mathbb{Z}$ and a prime number p. The Unique Factorization Theorem tells us that $n = \mathrm{sgn}(n)p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ for some prime numbers $p_1, \ldots, p_r$ and positive integers $\alpha_1, \ldots, \alpha_r$. Then we have

$$\left(\frac{n}{p}\right) = \left(\frac{\mathrm{sgn}(n)}{p}\right) \prod_{i=1}^{r} \left(\frac{p_i}{p}\right)^{\alpha_i}.$$

Thus all we need to do is to compute $\left(\frac{-1}{p}\right)$, $\left(\frac{2}{p}\right)$ and $\left(\frac{l}{p}\right)$ for any odd prime number l.

We have discussed $\left(\frac{-1}{p}\right)$ in Proposition 13.2. Now let us compute $\left(\frac{2}{p}\right)$. Again we may assume $p > 2$. Unlike previously, we do not have an information on whether $2^{\frac{p-1}{2}}$ is 1 modulo p. The new idea is to try finding a square root of 2 in $\overline{\mathbb{F}}_p$ and check to see whether it is in $\mathbb{F}_p$. Recall that $\mathbb{F}_p$ is precisely the set of elements of $\overline{\mathbb{F}}_p$ that satisfies $x^p = x$. Finally using the fact that $(a+b)^p = a^p + b^p$ in $\overline{\mathbb{F}}_p$ will allow us to simplify the computation.

Proposition 13.5. *2 is a quadratic residue modulo p if and only if $p = 2$ or $p \equiv \pm 1 \pmod 8$.*

Proof. 2 is certainly a quadratic residue modulo 2. From now on we assume $p > 2$.

Since $p \neq 2$, there exists a primitive 8-th root of unity α in $\overline{\mathbb{F}}_p$ by Exercise 2 of Assignment 10/19.

Consider $x_0 = \alpha + \alpha^{-1} \in \overline{\mathbb{F}}_p$. Then

$$x_0^2 = \alpha^2 + \alpha^{-2} + 2 = \alpha^{-2}(1 + \alpha^4) + 2. \tag{13.2}$$

But α is primitive 8-th root of unity, so $\alpha^8 = 1$ but $\alpha^4 \neq 1$. From $\alpha^8 = 1$, we get $(1+\alpha^4)(1-\alpha^4) = 0$. But $1 - \alpha^4 \neq 0$, so $1 + \alpha^4 = 0$. Hence $x_0^2 = 2$ by (13.2). Therefore x_0 and $-x_0$ are roots of the polynomial $x^2 - 2 \in \overline{\mathbb{F}}_p[x]$. But $\deg(x^2 - 2) = 2$, so $x^2 - 2$ has 2 roots in $\overline{\mathbb{F}}_p$. Thus the polynomial $x^2 - 2$ has a root in $\mathbb{F}_p$ if and only if $x_0 \in \mathbb{F}_p$. In other words, we have proven

$\boxed{\textit{Claim}}$ 2 is a quadratic residue modulo p if and only if $x_0 \in \mathbb{F}_p$.

On the other hand $x_0^p = (\alpha + \alpha^{-1})^p = \alpha^p + \alpha^{-p}$. Since $\alpha^8 = 1$ and $\alpha^4 = -1$ (we proved this in last paragraph), we have $x_0^p = x_0$ if $p \equiv \pm 1 \pmod 8$ and $x_0^p = -x_0$ if $p \equiv \pm 3 \pmod 8$.

Recall $\mathbb{F}_p = \{x \in \overline{\mathbb{F}}_p : x^p = x\}$. Since $p > 2$, we have $x_0 \neq -x_0$ in $\overline{\mathbb{F}}_p$. So by the previous paragraph, we have $x_0 \in \mathbb{F}_p \Leftrightarrow p \equiv \pm 1 \pmod 8$. Now combined with the claim above, we can conclude. □

Before going on, let us point out the two important steps in this proof ($l = 2$ in the rest of the paragraph): (1) find a square root, say x_0, of l or $-l$ in $\overline{\mathbb{F}}_p$ (the upshot is that $x_0 \in \mathbb{F}_p$ if and only if l (or $-l$) is a quadratic residue modulo p) such that x_0^p is easy to compute; (2) see whether x_0 is in $\mathbb{F}_p$ by using the fact $\mathbb{F}_p = \{x \in \overline{\mathbb{F}}_p : x^p = x\}$.

13.2. **Law of Quadratic Reciprocity.** One of Gauß's most amazing contribution to mathematics is his law of quadratic reciprocity.

Theorem 13.6. *Suppose p and l are two distinct odd prime numbers. Then*

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

In particular,

$$\left(\frac{p}{l}\right) = -\left(\frac{l}{p}\right) \text{ if } p, l \equiv 3 \pmod 4;$$
$$\left(\frac{p}{l}\right) = \left(\frac{l}{p}\right) \text{ otherwise.}$$

For example if we want to compute $\left(\frac{11}{29}\right)$, then we can do

$$\left(\frac{11}{29}\right) = \left(\frac{29}{11}\right) = \left(\frac{7}{11}\right) = -\left(\frac{11}{7}\right) = -\left(\frac{4}{7}\right) = -1.$$

The proof of Theorem 13.6 follows the same framework as underlined in the last paragraph of the previous subsection. What gets (much) more complicated is the computation.

14. Quadratic Reciprocity (continued)

Today we prove Gauß's law of quadratic reciprocity, namely Theorem 13.6.

Recall that when computing $\left(\frac{2}{p}\right)$ in Proposition 13.5, the two important steps are (with $l = 2$): (1) find a square root, say x_0, of l or $-l$ in $\overline{\mathbb{F}}_p$ (the upshot is that $x_0 \in \mathbb{F}_p$ if and only if l (or $-l$) is a quadratic residue modulo p) such that x_0^p is easy to compute; (2) see whether x_0 is in $\mathbb{F}_p$ by using the fact $\mathbb{F}_p = \{x \in \overline{\mathbb{F}}_p : x^p = x\}$.

The proof of Theorem 13.6 follows the same framework. What gets (much) more complicated is the computation.

Proof of Theorem 13.6. $\boxed{\textit{Step 1}}$ Find a square root, say x_0, of l or $-l$ in $\overline{\mathbb{F}}_p$ such that x_0^p is easy to compute.

Let α be a primitive l-th root of unity in $\overline{\mathbb{F}}_p$, namely in the group $\overline{\mathbb{F}}_p^\times$ we have $\mathrm{ord}(\alpha) = l$. See Exercise 2 of Assignment 10/19 for the existence of such an α.

For any congruence class $[m] \in \mathbb{F}_l$, define $\alpha^{[m]}$ to be $\alpha^m \in \overline{\mathbb{F}}_p$. Note that $\alpha^{[m]}$ is well-defined: if $[m] = [m']$ in $\mathbb{F}_l = \mathbb{Z}/l\mathbb{Z}$, then $\alpha^m = \alpha^{m'}$ since $\alpha^l = 1$.

From now on we will simplify the notation (drop the "[]") for elements of $\mathbb{F}_l$. Consider the Gauß sum

$$x_0 = \sum_{m \in \mathbb{F}_l} \left(\frac{m}{l}\right) \alpha^m \in \overline{\mathbb{F}}_p.$$

We claim that

$$x_0^2 = \left(\frac{-1}{l}\right) l. \tag{14.1}$$

Indeed, squaring x_0 gives

$$\begin{aligned} x_0^2 &= \sum_{m \in \mathbb{F}_l} \sum_{n \in \mathbb{F}_l} \left(\frac{mn}{l}\right) \alpha^{m+n} \\ &= \sum_{m \in \mathbb{F}_l} \sum_{t \in \mathbb{F}_l} \left(\frac{m(t-m)}{l}\right) \alpha^t \end{aligned}$$

where we made a change of variable, setting t equal to $m + n$. When $m = 0$, the Legendre symbol $\left(\frac{m(t-m)}{l}\right)$ is 0. When $m \neq 0$, we have

$$\left(\frac{m(t-m)}{l}\right) = \left(\frac{-m^2(1-tm^{-1})}{l}\right) = \left(\frac{-1}{l}\right)\left(\frac{m^2}{l}\right)\left(\frac{1-tm^{-1}}{l}\right) = \left(\frac{-1}{l}\right)\left(\frac{1-tm^{-1}}{l}\right).$$

Combing with equality above gives

$$\begin{aligned} x_0^2 &= \sum_{m \in \mathbb{F}_l^\times} \sum_{t \in \mathbb{F}_l} \left(\frac{-1}{l}\right)\left(\frac{1-tm^{-1}}{l}\right) \alpha^t \\ &= \left(\frac{-1}{l}\right) \sum_{t \in \mathbb{F}_l} \left(\sum_{m \in \mathbb{F}_l^\times} \left(\frac{1-tm^{-1}}{l}\right) \right) \alpha^t. \end{aligned} \tag{14.2}$$

When $t = 0$, we have

$$\sum_{m\in\mathbb{F}_l^\times}\left(\frac{1-tm^{-1}}{l}\right) = \sum_{m\in\mathbb{F}_l^\times}\left(\frac{1}{l}\right) = \sum_{m\in\mathbb{F}_l^\times} 1 = l-1.$$

When $t \neq 0$, we see that as m runs over every element of $\mathbb{F}_l^\times$, $1-tm^{-1}$ runs over every element of $\mathbb{F}_l \setminus \{1\}$. Hence

$$\sum_{m\in\mathbb{F}_l^\times}\left(\frac{1-tm^{-1}}{l}\right) = \sum_{s\in\mathbb{F}_l}\left(\frac{s}{l}\right) - \left(\frac{1}{l}\right) = \sum_{s\in\mathbb{F}_l^\times}\left(\frac{s}{l}\right) + \left(\frac{0}{l}\right) - \left(\frac{1}{l}\right) = \sum_{s\in\mathbb{F}_l^\times}\left(\frac{s}{l}\right) - 1.$$

In Exercise 1 of Assignment 10/24, we have proven that half of the elements of $\mathbb{F}_l^\times$ are squares in $\mathbb{F}_l = \mathbb{Z}/l\mathbb{Z}$ and half of the elements of $\mathbb{F}_l^\times$ are not squares in $\mathbb{F}_l = \mathbb{Z}/l\mathbb{Z}$. Hence $\left(\frac{s}{l}\right)$ equals 1 for half of $s \in \mathbb{F}_l^\times$ and equals -1 for half of $s \in \mathbb{F}_l^\times$. Hence $\sum_{s\in\mathbb{F}_l^\times}\left(\frac{s}{l}\right) = 0$. Therefore we have

$$\sum_{m\in\mathbb{F}_l^\times}\left(\frac{1-tm^{-1}}{l}\right) = -1$$

for any $t \neq 0$. Hence we get by (14.2)

$$\begin{aligned}
x_0^2 &= \left(\frac{-1}{l}\right)\left(l-1+\sum_{t\in\mathbb{F}_l^\times}(-\alpha^t)\right)\\
&= \left(\frac{-1}{l}\right)\left(l-1-\sum_{t=1}^{l-1}\alpha^t\right)\\
&= \left(\frac{-1}{l}\right)(l-1-\frac{\alpha(1-\alpha^{l-1})}{1-\alpha})\\
&= \left(\frac{-1}{l}\right)(l-1-\frac{\alpha-\alpha^l}{1-\alpha})\\
&= \left(\frac{-1}{l}\right)(l-1-\frac{\alpha-1}{1-\alpha})\\
&= \left(\frac{-1}{l}\right)l,
\end{aligned}$$

which is precisely what we desire (14.1).

Let us see the upshot of Step 1. Now x_0 and $-x_0$ are roots of the polynomial $x^2 - \left(\frac{-1}{l}\right)l \in \overline{\mathbb{F}}_p[x]$. But $\deg(x^2 - \left(\frac{-1}{l}\right)l) = 2$, so $x^2 - \left(\frac{-1}{l}\right)l$ has 2 roots in $\overline{\mathbb{F}}_p$. Thus the polynomial $x^2 - \left(\frac{-1}{l}\right)l$ has a root in $\mathbb{F}_p$ if and only if $x_0 \in \mathbb{F}_p$. Hence we have

$$\left(\frac{-1}{l}\right)l \text{ is a quadratic residue modulo } p \text{ if and only if } x_0 \in \mathbb{F}_p.$$

Since $\gcd(l,p) = 1$, the sentence above becomes

$$\left(\frac{\left(\frac{-1}{l}\right)l}{p}\right) = 1 \Leftrightarrow x_0 \in \mathbb{F}_p.$$

By Proposition 13.3 (applied to $l > 2$ and then to $p > 2$), we have

$$\left(\frac{\left(\frac{-1}{l}\right)l}{p}\right) = \left(\frac{(-1)^{\frac{l-1}{2}}l}{p}\right) = \left(\frac{(-1)^{\frac{l-1}{2}}}{p}\right)\left(\frac{l}{p}\right) = (-1)^{\frac{l-1}{2}\frac{p-1}{2}}\left(\frac{l}{p}\right),$$

so

$$(-1)^{\frac{l-1}{2}\frac{p-1}{2}}\left(\frac{l}{p}\right) = 1 \Leftrightarrow x_0 \in \mathbb{F}_p.$$

Step 2 See whether x_0 is in $\mathbb{F}_p$ by using the fact $\mathbb{F}_p = \{x \in \overline{\mathbb{F}}_p : x^p = x\}$.

Since $\mathrm{char}(\overline{\mathbb{F}}_p) = p$, we have by Exercise 3 of Assignment 10/17

$$x_0^p = \sum_{m \in \mathbb{F}_l} \left(\frac{m}{l}\right)^p \alpha^{mp}.$$

By definition of the Legendre symbol, $\left(\frac{m}{l}\right)$ can only be -1, 0 or 1 and all three of them raised to an odd power equal to themselves, we have

$$\begin{aligned} x_0^p &= \sum_{m \in \mathbb{F}_l} \left(\frac{m}{l}\right)\alpha^{mp} \\ &= \sum_{n \in \mathbb{F}_l} \left(\frac{np^{-1}}{l}\right)\alpha^{n} \text{ by change of variable by setting } n = mp \\ &= \left(\frac{p^{-1}}{l}\right)\sum_{n \in \mathbb{F}_l} \left(\frac{n}{l}\right)\alpha^{n} \\ &= \left(\frac{p}{l}\right)x_0. \end{aligned}$$

Thus

$$x_0^p = x_0 \Leftrightarrow \left(\frac{p}{l}\right) = 1,$$

and hence

$$x_0 \in \mathbb{F}_p \Leftrightarrow \left(\frac{p}{l}\right) = 1.$$

Now let us look at the conclusions of the two steps. By definition, the Legendre symbol can only be -1, 0 or 1. But $\gcd(p, l) = 1$, so $\left(\frac{l}{p}\right) \neq 0$ and $\left(\frac{p}{l}\right) \neq 0$. Then the conclusions of the two steps imply

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

Multiplying both sides by $\left(\frac{p}{l}\right)$, we have

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

Hence we are done. □

15. Arithmetic in $\mathbb{Z}[i]$

15.1. **Sum of two squares.** We now work towards the following result which has been promised a long time ago. This gives the integer solutions to $z = x^2 + y^2$.

Theorem 15.1. *An integer $m \geq 1$ is expressible as a sum of two squares if and only if $m = 2^\alpha p_1^{\alpha_1} \cdots p_n^{\alpha_n}$ with α_i even when $p_i \equiv 3 \pmod 4$.*

Proof of $\Rightarrow$. Let us prove the necessity, namely $\Rightarrow$, first. The statement then becomes: Let p be an odd prime number such that $p \equiv 3 \pmod 4$. Let $m \geq 1$ be an integer. If $m = x^2 + y^2$ for some $x, y \in \mathbb{Z}$, then the greatest integer γ such that $p^\gamma | m$ is even.

Let us prove this statement by induction on m. Note that $m = x^2 + y^2$ and $p|m$ together imply that $x^2 + y^2 \equiv 0 \pmod p$.
(base step) $\mathcal{P}(1)$ is clearly true as $\gamma = 0$ is even in this case.
(induction step) Assume $\mathcal{P}(1), \ldots, \mathcal{P}(m-1)$ are true.

Assume $p \nmid x$. Then x is invertible in $\mathbb{F}_p$. Then we have $1 + (yx^{-1})^2 = 0$ in $\mathbb{F}_p$. In other words, $-1 = (yx^{-1})^2$ is a square in $\mathbb{F}_p$. Thus $\left(\frac{-1}{p}\right) = 1$. But then Proposition 13.2 implies $p \equiv 1 \pmod 4$, contradicting the assumption $p \equiv 3 \pmod 4$.

Hence $p|x$. Then $x^2 + y^2 \equiv 0 \pmod p$ implies that $p|y$. But then $p^2 | x^2 + y^2 = m$. Now $m/p^2 \in \mathbb{Z}$ is the sum of two squares $(x/p)^2 + (y/p)^2$.

Since $m/p^2 < m$, we can apply induction hypothesis to conclude that $\gamma - 2$ is even. So γ is even. □

In the proof, we see that the equation $x^2 + y^2 = 0$ in $\mathbb{F}_p$ only has $(0,0)$ as solutions when $p \equiv 3 \pmod 4$. We call $(0,0)$ the **trivial solution** to this equation. Similarly we say that a solution (x_0, y_0) to $x^2 + y^2 \equiv 0 \pmod p$ is trivial if $p|x_0$ and $p|y_0$.

Proving the sufficiency $\Leftarrow$ of Theorem 15.1 is more involved. We start with the following observation, whose motivation we will see later.

$$(x^2 + y^2)(u^2 + v^2) = (xu + yv)^2 + (xv - yu)^2. \tag{15.1}$$

In other words, the product of two integers both expressible as sums of two squares is also expressible as a sum of two squares. Since $2 = 1^2 + 1^2$ and p^2 are both expressible as a sum of two squares, it suffices to prove the following statement.

Theorem 15.2. *Let p be an odd prime number such that $p \equiv 1 \pmod 4$. Then p is expressible as a sum of two squares.*

15.2. **The ring $\mathbb{Z}[i]$ of Gaussian integers.** Let $\overline{\mathbb{Q}}$ be an algebraic closed field containing $\mathbb{Q}$. Let i be an element of $\overline{\mathbb{Q}}$ that satisfies $i^2 + 1 = 0$. Let $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\} \subset \overline{\mathbb{Q}}$. The addition and multiplication on $\overline{\mathbb{Q}}$ restrict to $\mathbb{Z}[i]$, or equivalently $\mathbb{Z}[i]$ is closed under addition and multiplication: The laws are defined as follows

$$\begin{aligned}(a + bi) + (c + di) &= (a + c) + (b + d)i;\\ (a + bi)(c + di) &= (ac - bd) + (ad + bc)i.\end{aligned}$$

So $(\mathbb{Z}[i], +, \cdot; 0, 1)$ is a commutative ring.

We want to compute $\mathbb{Z}[i]^\times = \{\alpha \in \mathbb{Z}[i] : \alpha\beta = 1 \text{ for some } \beta \in \mathbb{Z}[i]\}$. As for $\mathbb{Z}$ and $F[x]$, it will be more convenient to do the computation if there is a size function on $\mathbb{Z}[i]$. Fortunately such a nice function does exist for $\mathbb{Z}[i]$. Let us define it now.

For any $\alpha = a + bi \in \mathbb{Z}[i]$, we define its **norm** by $N(\alpha) = (a + bi)(a - bi) = a^2 + b^2$. Then $N(\alpha) \in \mathbb{Z}$, $N(\alpha) \geq 0$ and $N(\alpha) = 0 \Leftrightarrow \alpha = 0$. Moreover (15.1) implies that

$$N(\alpha\beta) = N(\alpha)N(\beta).$$

It is not hard to see that $N(\alpha) \geq 1$ for any $\alpha \neq 0$.

With this size function, we can compute $\mathbb{Z}[i]^\times$ as for $\mathbb{Z}$ and $F[x]$.

Proposition 15.3. $\mathbb{Z}[i]^\times = \{1, -1, i, -i\}$.

Proof. It is clear that $Z[i]^\times \supset \{1, -1, i, -i\}$. Let us prove the other inclusion.

Suppose $\alpha \in \mathbb{Z}[i]^\times$. Then $\alpha\beta = 1$ for some $\beta \in \mathbb{Z}[i]$. In particular $\alpha, \beta \neq 0$. Taking the norms of both sides, we have $N(\alpha)N(\beta) = N(\alpha\beta) = N(1) = 1$. But $N(\alpha), N(\beta) \geq 1$, so we must have $N(\alpha) = N(\beta) = 1$. Write $\alpha = a + bi$. Then $a^2 + b^2 = N(\alpha) = 1$. But then $a = \pm 1, b = 0$ or $a = 0, b = \pm 1$. So $\alpha = \pm 1, \pm i$. Hence $\mathbb{Z}[i]^\times \subset \{1, -1, i, -i\}$. Hence we are done. □

Note that we showed in the proof $N(\alpha) = 1 \Leftrightarrow \alpha = \pm 1, \pm i \Leftrightarrow \alpha \in \mathbb{Z}[i]^\times$.

As an exercise, prove the following lemma. (Compare this with Lemma 7.5.)

Lemma 15.4. *If $\alpha\mathbb{Z}[i] = \beta\mathbb{Z}[i]$, then $\alpha \in \beta\mathbb{Z}[i]^\times$, namely $\alpha = \beta\gamma$ for some $\gamma \in \{\pm 1, \pm i\}$.*

We also have the following division algorithm in $\mathbb{Z}[i]$.

Proposition 15.5 (Division Algorithm)**.** *For any $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$, there exist elements $\gamma, \delta \in \mathbb{Z}[i]$ such that*

$$\alpha = \beta\gamma + \delta, \quad \text{with } N(\delta) < N(\beta).$$

Note that (γ, δ) may not be unique. For example

$$i = (2 + i)0 + i,$$
$$i = (2 + i)1 + (-2),$$

but $N(i), N(-2) < N(2 + i)$. However in most applications of the division algorithm, the uniqueness is not important as we only want to find a "smaller" δ and do descent argument. Let us call this δ the **remainder**.

Proof of Proposition 15.5. Write $\alpha = a + bi$ and $\beta = c + di$. Since $\alpha, \beta \in \overline{\mathbb{Q}}$, we may compute $\alpha\beta^{-1}$ in $\overline{\mathbb{Q}}$ even though β^{-1} may not be in $\mathbb{Z}[i]$:

$$\frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)} = \frac{(ac + bd) + (bc - ad)i}{c^2 + d^2} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i =: s + ti.$$

Now s, t are rational numbers. Let x, y be the closest integers to s, t respectively, namely $x, y \in \mathbb{Z}$ such that $|x - s| \leq \frac{1}{2}$ and $|y - t| \leq \frac{1}{2}$.[6] Set $\gamma = x + yi$ and $\delta = \alpha - \beta\gamma$. Then it suffices to prove $N(\delta) < N(\beta)$. But now $\delta = \beta((s - x) + (t - y)i)$, so

$$N(\delta) = N(\beta)N((s - x) + (t - y)i) = N(\beta)(|s - x|^2 + |t - y|^2) \leq N(\beta)(\frac{1}{4} + \frac{1}{4}) < N(\beta).$$

□

[6] x and y may not be unique, for example when $s = 3.5$. But here we only need one such x and y.

We illustrate this algorithm with an example: take $\alpha = 1 - 8i$ and $\beta = 2 + i$, then

$$\frac{1-8i}{2+i} = \frac{(1-8i)(2-i)}{(2+i)(2-i)} = \frac{-6-17i}{5} = -1.2 - 3.4i.$$

Then we take $x = -1$ and $y = -3$. Hence set $\gamma = -1 - 3i$ and $\delta = \alpha - \beta\gamma = -i$. Sure enough $N(-i) < N(2+i)$.

One major advantage of the division algorithm, as we have seen twice in this course, is that it allows us to have the Euclid's Algorithm in $\mathbb{Z}[i]$. As before, we define $\gcd(\alpha, \beta) \in \mathbb{Z}[i]$ to be such that $\alpha\mathbb{Z}[i] + \beta\mathbb{Z}[i] = \gcd(\alpha, \beta)\mathbb{Z}[i]$. We know that $\gcd(\alpha, \beta)$ is well-defined up to $\mathbb{Z}[i]^\times$ by Lemma 15.4. A candidate for $\gcd(\alpha, \beta)$ is the last non-zero remainder in the Euclid's Algorithm applied to α and β.

Now we are ready to prove Theorem 15.2.

Proof of Theorem 15.2. Let p be a prime number such that $p \equiv 1 \pmod 4$. Then $x^2 + 1 \equiv 0 \pmod p$ has solution by Proposition 13.2. In other words there exists an integer a such that $p | a^2 + 1 = N(a + i)$. Set $\alpha = \gcd(p, a + i)$. Then $\alpha\mathbb{Z}[i] = p\mathbb{Z}[i] + (a + i)\mathbb{Z}[i] \ni p$, and hence $p = \alpha\beta$ for some $\beta \in \mathbb{Z}[i]$. So $N(\alpha)|N(p) = p^2$ in $\mathbb{Z}$.[7] There are three choices for $N(\alpha)$: 1, p and p^2.

If $N(\alpha) = 1$, then we have seen that $\alpha = \pm 1, \pm i$. Then $p\mathbb{Z}[i] + (a + i)\mathbb{Z}[i] = \alpha\mathbb{Z}[i] = \mathbb{Z}[i]$. So $p\beta + (a+i)\gamma = 1$ for some $\beta, \gamma \in \mathbb{Z}[i]$. Write $p\beta = c + di$, then $p|c$, $p|d$ and $(a+i)\gamma = (1-c) - di$. Then

$$N(a + i)N(\gamma) = N((1 - c) - di) = (1 - c)^2 + d^2 \equiv 1 \pmod p.$$

But we have seen $p|N(a + i)$, so $N(a + i)N(\gamma) \equiv 0 \pmod p$. Contradiction.

If $N(\alpha) = p^2$. Recall $p = \alpha\beta$. Taking the norms on both sides we get $N(\beta) = 1$. So $\beta = \pm 1, \pm i$ and hence $\alpha = \pm pi, \pm p$. However $\alpha | a + i$ in $\mathbb{Z}[i]$, so $p | a + i$ in $\mathbb{Z}[i]$. Hence $a + i = p(c + di)$ for some $c, d \in \mathbb{Z}$. But then $1 = pd$, which is impossible.

So we must have $N(\alpha) = p$. Writing $\alpha = k + li$ then yields $p = k^2 + l^2$ being a sum of two squares. □

[7] Taking the complex conjugation for $p = \alpha\beta$, we get $p = \overline{\alpha}\overline{\beta}$. Then $N(p) = p^2 = (\alpha\overline{\alpha})(\beta\overline{\beta}) = N(\alpha)N(\beta)$.

16. Arithmetic in $\mathbb{Z}[i]$ (continued)

16.1. **Guassian primes.** Just as in the cases of $\mathbb{Z}$ and $F[x]$, we wish to know what the analogous notion of "prime" is for $\mathbb{Z}[i]$. Here are two equivalent definitions.

Definition 16.1. *A nonzero element α of $\mathbb{Z}[i]$ is a **prime** if $\alpha \notin \mathbb{Z}[i]^\times$ and any divisor of α belongs to $\mathbb{Z}[i]^\times \bigcup \alpha\mathbb{Z}[i]^\times$.*

Definition 16.2. *A nonzero element α of $\mathbb{Z}[i]$ is a **prime** if $\alpha \notin \mathbb{Z}[i]^\times$ and $\alpha|\beta\gamma \Rightarrow \alpha|\beta$ or $\alpha|\gamma$ $(\forall \beta, \gamma \in \mathbb{Z}[i])$.*

To distinguish between a prime in $\mathbb{Z}[i]$ with a prime in $\mathbb{Z}$, we call a prime in $\mathbb{Z}[i]$ a **Gaussian prime** and the latter a **prime number**.

Again as for $\mathbb{Z}$ and $F[x]$, the proof of Definition 16.2⇒Definition 16.1 is straight-forward, and Definition 16.1⇒Definition 16.2 follows from the notion of gcd. Moreover, we also have the following Unique Factorization Theorem.

Theorem 16.3. *Let $\alpha \in \mathbb{Z}[i]$ such that $\alpha \neq 0$ and $\alpha \notin \mathbb{Z}[i]^\times$. Then α can be factorized into a product of Gaussian primes, unique up to reordering and $\mathbb{Z}[i]^\times$. More precisely, we have*

(1) $\alpha = \mathfrak{p}_1 \cdots \mathfrak{p}_n$ for some Gaussian primes $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$.
(2) If $\alpha = \mathfrak{p}_1 \cdots \mathfrak{p}_n = \mathfrak{p}'_1 \cdots \mathfrak{p}'_m$ for Gaussian primes $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ and $\mathfrak{p}'_1, \ldots, \mathfrak{p}'_m$, then $n = m$ and we can reorder $\mathfrak{p}'_1, \ldots, \mathfrak{p}'_n$ such that $\mathfrak{p}_i \in \mathfrak{p}'_i\mathbb{Z}[i]^\times$ for all $i \in \{1, \ldots, n\}$.

Next, let us give a classification of Gaussian primes.

Proposition 16.4. *If α is a Gaussian prime, then α divides a prime number.*

Proof. Write $\alpha = a + bi$. Then $N(\alpha) = (a+bi)(a-bi) = \alpha(a-bi)$ is a positive integer divisible by α. Factorize $N(\alpha)$ into a product of prime numbers $p_1 \cdots p_n$ in $\mathbb{Z}$, then $\alpha|N(\alpha) = p_1 \cdots p_n$. By Definition 16.2, we have $\alpha|p_i$ for some $i \in \{1, \ldots, n\}$. Hence we are done. □

Proposition 16.5. *Suppose $\alpha \in \mathbb{Z}[i]$ and $N(\alpha)$ is a prime number. Then α is a Gaussian prime.*

Proof. Clearly $\alpha \neq 0$ and $\alpha \notin \mathbb{Z}[i]^\times$. We use Definition 16.1. Suppose $\alpha = \beta\gamma$ for $\beta, \gamma \in \mathbb{Z}[i]$. We need to show that either $\beta \in \mathbb{Z}[i]^\times$ or $\gamma \in \mathbb{Z}[i]^\times$.

Taking norms gives $N(\alpha) = N(\beta)N(\gamma)$. Since $N(\alpha)$ is a prime number, either $N(\beta)$ or $N(\gamma)$ is 1. Then the conclusion follows from the paragraph below Proposition 15.3. □

Corollary 16.6. *Let p be a prime number. Suppose $p = x^2 + y^2 = z^2 + w^2$ for non-negative $x, y, z, w \in \mathbb{Z}$. Then $\{x, y\} = \{z, w\}$.*

Proof. Let $\alpha = x + yi$, $\alpha' = x - yi$ and $\beta = z + wi$, $\beta' = z - wi$. Then $p = N(\alpha) = N(\alpha') = N(\beta) = N(\beta')$. By Proposition 16.5, we know that α, α' and β, β' are Gaussian primes. Thus we have two factorizations of $p \in \mathbb{Z}[i]$ into the product of Gaussian primes: $p = \alpha\alpha'$ and $p = \beta\beta'$. Then the uniqueness part of UFT for $\mathbb{Z}[i]$, namely Theorem 16.3.(2), implies the result (recall that $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$). □

This corollary supplements Theorem 15.2 and Theorem 15.1. Let us write this more clearly.

Theorem 16.7. *A prime number p is expressible as a sum of two squares if and only if $p = 2$ or $p \equiv 1 \pmod 4$. Moreover when it happens, then the expression is unique.*

Proposition 16.8. *Suppose p is a prime number such that $p \equiv 3 \pmod 4$. Then p is a Gaussian prime.*

Proof. Clearly $\alpha \neq 0$ and $\alpha \notin \mathbb{Z}[i]^\times$. We use Definition 16.1. Suppose $p = \beta\gamma$ for some $\beta, \gamma \in \mathbb{Z}[i]$. Taking norms gives $p^2 = N(\beta)N(\gamma)$. Since p is not a sum of two squares by Theorem 15.1, p is not the norm of any element of $\mathbb{Z}[i]$. Hence either $N(\beta)$ or $N(\gamma)$ is 1. Equivalently either $\beta \in \mathbb{Z}[i]^\times$ or $\gamma \in \mathbb{Z}[i]^\times$. Thus p is a Gaussian prime. □

Now we can summarize the discussion above and give the classification of Gaussian primes.

Theorem 16.9. *A Gaussian prime α is, up to $\mathbb{Z}[i]^\times$,*

(1) $1 + i$, or
(2) prime numbers congruent to 3 modulo 4, or
(3) elements β such that $N(\beta)$ is a prime number congruent to 1 modulo 4.

Proof. By Proposition 16.5 (applied to case 1 and 3) and Proposition 16.8 (applied to case 2), we know that every element in the list above is a Gaussian prime.

Conversely given any Gaussian prime $\alpha \in \mathbb{Z}[i]$, it divides a prime number p by Proposition 16.4. So $p = \alpha\beta$ for some $\beta \in \mathbb{Z}[i]$.

If $p = 2$, then as $2 = -i(1+i)^2$ is a prime factorization of 2 in $\mathbb{Z}[i]$, we have that $\alpha \in (1+i)\mathbb{Z}[i]^\times$ by the uniqueness part of UFT for $\mathbb{Z}[i]$ (Theorem 16.3.(2)). This is case 1 in the list.

If $p \equiv 3 \pmod 4$, then p is a Gaussian prime by Proposition 16.8. Hence $\alpha \in p\mathbb{Z}[i]^\times$ by the uniqueness part of UFT for $\mathbb{Z}[i]$ (Theorem 16.3.(2)). This is case 2 in the list.

If $p \equiv 1 \pmod 4$, then $p = x^2 + y^2$ for some $x, y \in \mathbb{Z}$ by Theorem 15.2. But then $p = (x + yi)(x - yi)$. Now $N(x + yi) = N(x - yi) = p$ is a prime number, so both $x + yi$ and $x - yi$ are Gaussian primes by Proposition 16.5. So $p = (x + yi)(x - yi)$ is a factorization of p into a product of Gaussian primes. So $\alpha \in (x + yi)\mathbb{Z}[i]^\times$ by the uniqueness part of UFT for $\mathbb{Z}[i]$ (Theorem 16.3.(2)). This is case 3 in the list. □

16.2. **Factorizing Gaussian integers.** We know that Gaussian integers have unique factorization. It will be nice if we know how to explicitly factor a Gaussian integer into Gaussian primes. We illustrate this with an example. Let $\beta = 1 - 8i$. A Gaussian prime α dividing β must also divide

$$N(\beta) = 1 + 8^2 = 65 = 5 \times 13 = (2 + i)(2 - i)(3 + 2i)(3 - 2i).$$

Here since 5 and 13 are congruent to 1 modulo 4, we can decompose them as products of two Gaussian primes. The above equation also implies that α must be, up to $\mathbb{Z}[i]^\times$, one of $2 + i$, $2 - i$, $3 + 2i$, $3 - 2i$. We can simply divide β by all of them using the division algorithm to see which of them divides β:

$$\frac{1 - 8i}{2 + i} = \frac{(1 - 8i)(2 - i)}{(2 + i)(2 - i)} = \frac{-6 - 17i}{5} \notin \mathbb{Z}[i]$$
$$\frac{1 - 8i}{2 - i} = \frac{(1 - 8i)(2 + i)}{(2 - i)(2 + i)} = \frac{10 - 15i}{5} = 2 - 3i \in \mathbb{Z}[i]$$

Since $N(2 - 3i) = 2^2 + 3^2 = 13$ is a prime number, we know that $2 - 3i$ is a Gaussian prime. So $1 - 8i = (2 - i)(2 - 3i)$ is a factorization of $1 - 8i$ into a product of Gaussian primes. If we test for divisibility for the other two primes, then we will find $1 + 8i = (3 - 2i)(-1 + 2i)$.

Note that the two factorizations $(2-i)(2-3i) = (3-2i)(-1+2i)$ are essentially the same factorization: $2-i = -i(-1+2i)$ and $2-3i = -i(3-2i)$.

Let us summarize the factorization algorithm: For any Gaussian integer β, we should

(1) Compute $N(\beta)$ and factor it as a product $p_1^{\alpha_1}\cdots p_n^{\alpha_n}$ of prime numbers;
(2) For any $p_j = 2$, we have $(1+i)^{\alpha_j}|\beta$;
(3) For any $p_j \equiv 3 \pmod 4$, α_j must be even and $p_j^{\alpha_j/2}|\beta$;
(4) For any $p_j \equiv 1 \pmod 4$, write it as a product $p_j = (a+bi)(a-bi)$ of two Gaussian primes. Test which of these two Gaussian primes divides β. Divide β by the one that does divide β and repeat the process.

16.3. **Extra: Irreducible v.s. Prime.** We have studied three rings extensively in the class: $\mathbb{Z}$, $F[x]$ and $\mathbb{Z}[i]$. For all three, we defined prime elements in two equivalent ways. In general we have the following definitions.

Setting Let $(R, +, \cdot; 0_R, 1_R)$ be a commutative ring. In the discussion below we only consider the case where R is an integral domain, namely $ab = 0_R \Rightarrow a = 0_R$ or $b = 0_R$ $(\forall a, b \in R)$.

Definition 16.10. *Let R be an integral domain. An element $\alpha \in R$ is an **irreducible** element if $\alpha = \beta\gamma \Rightarrow \beta \in R^\times$ or $\gamma \in R^\times$ $(\forall \beta, \gamma \in R)$.*

Definition 16.11. *Let R be an integral domain. An element $\alpha \in R$ is a **prime** element if $\alpha|\beta\gamma \Rightarrow \alpha|\beta$ or $\alpha|\gamma$ $(\forall \beta, \gamma \in R)$.*

Proposition 16.12. *In an integral domain R, any prime element is irreducible.*

The proof is as before.

The converse is in general not true. Consider the ring $R = \mathbb{Z}[\sqrt{-5}]$ consisting of elements of the form $a + b\sqrt{-5}$. Then $2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$. Then 2 divides the product of $1+\sqrt{-5}$ and $1-\sqrt{-5}$ but it does NOT divide any one of them. So 2 is not a prime element. On the other hand, one defines the norm of an element of R by $N(a+b\sqrt{-5}) = a^2 + 5b^2$. Then similar to $\mathbb{Z}[i]$, we have $N(\alpha\beta) = N(\alpha)N(\beta)$ for any $\alpha, \beta \in R$. Then it is not hard to compute that no elements of 2 have norm 2. Hence 2 is an irreducible element since a divisor of 2 has norm either 1 or 4: in the former case it is in $\mathbb{Z}[\sqrt{-5}]^\times$, in the latter case it is 2 multiplied by an element of $\mathbb{Z}[\sqrt{-5}]^\times$.

The failure for an irreducible element to be a prime element is the lack of the notion of gcd. This already shows that we do not have Euclid's Algorithm, nor Division Algorithm in $\mathbb{Z}[\sqrt{-5}]$.

Definition 16.13. *We say that an integral domain R is a **principal ideal domain** if for any $\alpha, \beta \in R$, there exists an element $\gamma \in R$ such that $\alpha R + \beta R = \gamma R$.*

In particular, any integral domain with gcd is a principal ideal domain (the converse not true).

Proposition 16.14. *Suppose R is a principal ideal domain. Then every irreducible element is a prime element.*

Proof. Suppose $\alpha \in R$ is irreducible. Let $\beta, \gamma \in R$ be such that $\alpha|\beta\gamma$. Since R is a principal ideal domain, we have $\alpha R + \beta R = \delta R$ for some $\delta \in R$. Since $\alpha \in \alpha R + \beta R$, we have $\alpha = \delta\epsilon$ for some $\epsilon \in R$. Since α is irreducible, either δ or ϵ is in $R^\times$.

If $\delta \in R^\times$, then $\delta R = R$. So $\alpha x + \beta y = 1_R$ for some $x, y \in R$. Multiplying both sides by γ gives $\alpha x\gamma + \beta y\gamma = \gamma$. Since $\alpha|\beta\gamma$, we have $\alpha|\gamma$.

If $\epsilon \in R^\times$, then $\delta = \alpha\epsilon^{-1}$ is divisible by α. Since $\delta|\beta$, we get $\alpha|\beta$. □

17. Pythagorean triples and the descent

17.1. **The equation $x^2 + y^2 = z^2$.** This is the well-known Pythagorean theorem for the three sides of a right angle triangle. There are ancient greek tablets with several large solutions. It is quite likely they already have an algorithm to write down many solutions, perhaps all of them.

Observe that if a prime number p divides two of x, y, z, then it also divides the square of the third and hence also the third. If p divides all x, y, z, then replacing x, y, z by $x/p, y/p, z/p$ gives another set of solution. We say that a solution is **primitive** if x, y, z have no common prime factors. In this case, being primitive is equivalent to being pairwise coprime.

In order to get all solutions to $x^2 + y^2 = z^2$, it suffices to understand all primitive solutions. The following theorem gives all primitive solutions.

Theorem 17.1. *If (x_0, y_0, z_0) is a primitive solution to $x^2 + y^2 = z^2$, then one of x_0 and y_0 is odd and the other is even. Suppose y_0 is even without loss of generality, then there are coprime integers r and s such that*

$$\begin{aligned} x_0 &= r^2 - s^2, \\ y_0 &= 2rs, \\ z_0 &= r^2 + s^2. \end{aligned}$$

Proof. We first work modulo 4. Since only $0, 1$ are squares modulo 4, we see that it is impossible for x_0 and y_0 to be both odd: otherwise $z_0^2 = x_0^2 + y_0^2 \equiv 2 \pmod 4$ which cannot happen. So one of x_0 and y_0 is even. Since (x_0, y_0, z_0) is a primitive solution, we cannot have both x_0 and y_0 being even. So one of x_0 and y_0 is odd and the other is even.

Suppose y_0 is even without loss of generality. Then x_0 is odd, and hence z_0 is odd. We have

$$y_0^2 = z_0^2 - x_0^2 = (z_0 + x_0)(z_0 - x_0), \tag{17.1}$$

and hence

$$(\frac{y_0}{2})^2 = \frac{z_0 + x_0}{2} \frac{z_0 - x_0}{2}. \tag{17.2}$$

Note that $\frac{y_0}{2}$, $\frac{z_0+x_0}{2}$ and $\frac{z_0-x_0}{2}$ are integers because y_0 is even and x_0, z_0 are both odd.

We have $\gcd(z_0 + x_0, z_0 - x_0) = \gcd(z_0 + x_0, 2x_0)$. Since z_0 and x_0 are coprime, we conclude that $\gcd(z_0 + x_0, z_0 - x_0) = \gcd(z_0 + x_0, 2)$. But both x_0 and z_0 are odd, so $z_0 + x_0$ is even. So $\gcd(z_0 + x_0, z_0 - x_0) = 2$. So (17.2) is the factorization of the integer $(\frac{y_0}{2})^2$ into the product of two coprime integers. By UFT, both factors must be squares of integers, namely we have

$$\frac{z_0 + x_0}{2} = r^2 \text{ and } \frac{z_0 - x_0}{2} = s^2$$

for some $r, s \in \mathbb{Z}$ which are coprime to each other. So we get $x_0 = r^2 - s^2$ and $z_0 = r^2 + s^2$.

Now $(\frac{y_0}{2})^2 = r^2 s^2$, so replacing r by $-r$ if necessary, we have $y_0 = 2rs$. □

The key idea in the above proof is (17.1) where we express the power of an integer as a product of integers. Furthermore, we can compute the gcd of the factors so that the UFT then implies that each factor is not far from being a power themselves. We will use this technique a lot in the coming weeks!

17.2. **The equation $x^4 + y^4 = z^2$.** One of Fermat's most famous result is that the equation $x^4 + y^4 = z^2$ has no nontrivial solution. From this it follows that $x^4 + y^4 = z^4$ also has no nontrivial solution. Here a trivial solution means that one of the x, y, z is 0.

Theorem 17.2. *If (x_0, y_0, z_0) is a solution to $x^4 + y^4 = z^2$, then $x_0y_0z_0 = 0$.*

Proof. Assume none of x_0, y_0, z_0 is zero. By replacing x_0, y_0, z_0 by their negatives if necessary, we may assume they are all positive. Like before, we first reduce to the case where x_0, y_0, z_0 are pairwise coprime. If a prime number p divides two of them, then it also divides the third. If p divides x_0 and y_0, then p^4 divides z_0^2 and so p^2 divides z_0. Replacing (x_0, y_0, z_0) by $(x_0/p, y_0/p, z_0/p^2)$ gives another solution. Hence, we can once again only consider primitive solutions.

The key idea is to play with the following two equations

$$x^4 + y^4 = z^2 \tag{17.3}$$

and

$$x^2 + 4y^4 = z^4. \tag{17.4}$$

We will show: a primitive solution to (17.3) gives a solution to (17.4), and vice versa a primitive solution to (17.4) gives a solution to (17.3). However, if one starts with a primitive solution to (17.3), one ends up with another primitive solution to (17.3) but smaller. This process can be continued forever while positive integers cannot be decreased indefinitely while staying positive. Contradiction. This method is called Fermat's infinite descent.

Let us give more details of this descent process. Suppose (x_0, y_0, z_0) is a primitive solution to (17.3). Then (x_0^2, y_0^2, z_0) is a primitive solution to $x^2 + y^2 = z^2$. Without loss of generality we assume that y_0 is even. By Theorem 17.1, there are coprime integers r and s such that

$$\begin{aligned} x_0^2 &= r^2 - s^2, \\ y_0^2 &= 2rs, \\ z_0 &= r^2 + s^2. \end{aligned}$$

Since x_0 is odd, we see that $r^2 - s^2 = x_0^2 \equiv 1 \pmod 4$. So we must have $r^2 \equiv 1 \pmod 4$ and $s^2 \equiv 0 \pmod 4$ since any square is congruent to 0 or 1 modulo 4. Thus r is odd and s is even. The second equation then implies

$$(\frac{y_0}{2})^2 = r\frac{s}{2},$$

where r and $\frac{s}{2}$ are coprime (since r and s are already coprime). So UFT implies that $r = r_0^2$ and $s = 2s_0^2$ for some positive integers r_0 and s_0. Then the first equation shows that (x_0, s_0, r_0) is a solution to (17.4). Note that we have $z_0 \geq r^2 = r_0^4 > r_0$.

Now let us start with a solution (x_1, y_1, z_1) to (17.4). As before, if a prime number p divides two of x_1, y_1, z_1 then it also divides the third. If p divides y_1 and z_1, then $p^2 | x_1$ and $(x_1/p^2, y_1/p, z_1/p)$ also gives a solution to (17.4). Hence we may restrict ourselves to primitive solutions and assume that x_1, y_1, z_1 are pairwise coprime. So $(x_1, 2y_1^2, z_1^2)$ is a primitive solutioh to $x^2+y^2 = z^2$. Hence there are coprime positive integers r_1, s_1 such that

$$\begin{aligned} x_1 &= r_1^2 - s_1^2, \\ 2y_1^2 &= 2r_1s_1, \\ z_1^2 &= r_1^2 + s_1^2. \end{aligned}$$

By UFT, the second equation then gives $r_1 = r_2^2, s_1 = s_2^2$ for coprime positive integers r_2, s_2. Plugging them into the third equation gives $z_1^2 = r_2^4 + s_2^4$. Hence we get a solution (r_2, s_2, z_1) to (17.3).

Now we summarize our descent algorithm: We start with a positive primitive solution (x_0, y_0, z_0) to (17.3); we obtain a positive solution (x_1', y_1', z_1') to (17.4) with $z_1' < z_0$; we divide out by common factors to get a positive primitive solution (x_1, y_1, z_1) to (17.4) with $z_1 \leq z_1' < z_0$; we then obtain a positive solution (x_2', y_2', z_2') to (17.3) with $z_2' = z_1 < z_0$; we divide out by common factors to get a positive primitive solution (x_2, y_2, z_2) to (17.3) with $z_2 \leq z_2' < z_0$. The component z keeps decreasing and is positive. From this we obtain our contradiction since our algorithm can be repeated forever. □

18. Diophantine equations via local methods

Some Diophantine equations are easy to tell to have no solutions. For example, the equation $x^2+y^2=-1$ has no integer solution because it does not even have any real solutions. Here are a few more examples where solutions do not exist modulo some positive integer m. We view these methods as local methods.

Proposition 18.1. *The equation $x^2+y^2=4z+3$ has no integer solutions.*

Proof. Working modulo 4, we see that the right hand side is congruent to 3 modulo 4. Squares modulo 4 can only be 0 or 1, so the left hand side can only can 0 or 1 or 2 modulo 4. This implies the conclusion. □

Proposition 18.2. *The equation $15x^2-7y^2=9$ has no integer solutions.*

Proof. We work modulo 3. If (x_0,y_0) is a solution, then $3|7y_0^2$ and hence $3|y_0$. Write $y_0=3y_1$, then we get $5x_0^2-21y_1^2=3$.

Working modulo 3 again, we get $3|5x_0^2$ and hence $3|x_0$. Write $x_0=3x_1$, then we have $15x_1^2-7y_1^2=1$.

Observe that squares modulo 3 can only be 0 or 1. So the left hand side is congruent to 0 or 2 modulo 3. But the right hand side is congruent to 1 modulo 3. So we get a contradiction. □

Proposition 18.3. *The equation $x^2+y^2=3(z^2+w^2)$ has no nontrivial integer solutions.*

Proof. If (x_0,y_0,z_0,w_0) is a nontrivial solution, then dividing out the common factors gives rise to another nontrivial solution. Hence we may assume that the only positive integer dividing at the same time x_0,y_0,z_0,w_0 is 1.

Modulo 3 we get $x_0^2+y_0^2\equiv 0 \pmod 3$. Since squares modulo 3 are either 0 or 1, we get that $3|x_0$ and $3|y_0$. Write $x_0=3x_1$ and $y_0=3y_1$. Then we have $3(x_1^2+y_1^2)=z_0^2+w_0^2$. Again modulo 3 we get $z_0^2+w_0^2\equiv 0 \pmod 3$. Since squares modulo 3 are either 0 or 1, we again get $3|z_0$ and $3|w_0$.

Now 3 divides all of x_0,y_0,z_0,w_0. Contradiction. □

In the previous examples, it suffices to specify a modulus m and argue modulo m. The next example is a case where it is not enough to modulo a specific m but many different numbers m.

Proposition 18.4. *The equation $y^2=x^3+7$ has no integer solutions.*

Proof. Suppose (x_0,y_0) is a solution. We have

$$y_0^2=x_0^3+7\equiv x_0^3-1 \pmod 4.$$

If x_0 is even, then $4|x_0^3$ and hence $y_0^2\equiv -1 \pmod 4$, which cannot happen. So x_0 is odd and hence $x_0^2\equiv 1 \pmod 4$. Then $y_0^2=x_0^3+7$ is even and hence y_0 is even. Then the congruent equation above becomes

$$0\equiv x_0-1 \pmod 4.$$

So $x_0\equiv 1 \pmod 4$. Next observe that

$$y_0^2+1=x_0^3+2^3=(x_0+2)(x_0^2-2x_0+4).$$

Let us consider the prime factors of both sides. If p is such a prime number, then $p|y_0^2+1$. So $y_0^2\equiv -1 \pmod p$, namely -1 is a quadratic residue modulo p. So $p=2$ or $p\equiv 1 \pmod 4$ by Proposition 13.2. But y_0^2+1 is odd, so $p\equiv 1 \pmod 4$.

Thus by UFT, x_0+2 is the product of some prime numbers which are congruent to 1 modulo 4. But then $x_0+2\equiv 1 \pmod 4$. This contradicts $x_0\equiv 1 \pmod 4$. □

There are also plenty of Diophantine equations have solutions in real numbers and modulo m or every positive integer m while having no solutions in $\mathbb{Z}$. We call this the failure of the **local-global principle**. Next we give an example where the local-global principle fails.

Proposition 18.5. *Let $f(x)=(x^2-17)(x^2-19)(x^2-17\times 19)$. Then $f(x)\equiv 0 \pmod m$ has a solution for any positive integer m, but $f(x)=0$ has no integer solutions.*

Proof. We know that the solutions to $f(x)=0$ are $\pm\sqrt{17}$, $\pm\sqrt{19}$ and $\pm\sqrt{17\times 19}$. None of them is an integer. So $f(x)=0$ has no integer solutions.

Next let us prove that $f(x)\equiv 0 \pmod m$ has a solution for any positive integer m. By Chinese Remainder Theorem, it suffices to prove this for all $m=p^r$ for some prime number p and some integer $r\geq 1$.

For any prime number p, we have

$$\left(\frac{17}{p}\right)\left(\frac{19}{p}\right)\left(\frac{17\times 19}{p}\right)=\left(\frac{17}{p}\right)^2\left(\frac{19}{p}\right)^2=0 \text{ or } 1.$$

If the product is 0, then $[0]$ is a solution to $f(x)\equiv 0 \pmod p$.

If the product is 1, then one of $\left(\frac{17}{p}\right)$, $\left(\frac{19}{p}\right)$ and $\left(\frac{17\times 19}{p}\right)$ is 1. Hence one of $[17]$, $[19]$ and $[17\times 19]$ is a square in $\mathbb{F}_p$. This gives a solution x_0 to $f(x)\equiv 0 \pmod p$.

When $p\neq 2,17,19$, then we can check that $f'(x_0)\not\equiv 0 \pmod p$. So by Hensel's Lemma, $f(x)\equiv 0 \pmod{p^r}$ has a solution for any $r\geq 1$.

When $p=2$, then 1 is a solution to $x^2\equiv 17 \pmod 8$. So by Exercise 3 of Assignment 10/24, there is a solution to $x^2\equiv 17 \pmod{2^r}$ for any $r\geq 3$. So $f(x)\equiv 0 \pmod{2^r}$ has a solution for any $r\geq 1$.

When $p=17$, then

$$\left(\frac{19}{p}\right)=\left(\frac{19}{17}\right)=\left(\frac{2}{17}\right)=1$$

where the last equality follows from Proposition 13.5. So 19 is a quadratic residue modulo 17. Now suppose $x_0^2\equiv 19 \pmod{17}$. Then $x_0\not\equiv 0 \pmod{17}$, and hence $f'(x_0)\equiv 4x_0^3\cdot 2x_0=8x_0^4\not\equiv 0 \pmod{17}$. So by Hensel's Lemma, x_0 lifts to a solution to $x^2-19\equiv 0 \pmod{17^r}$ for any $r\geq 1$. Hence $f(x)\equiv 0 \pmod{17^r}$ has a solution for any $r\geq 1$.

When $p=19$, then

$$\left(\frac{17}{p}\right)=\left(\frac{17}{19}\right)=\left(\frac{-2}{19}\right)=\left(\frac{-1}{19}\right)\left(\frac{2}{19}\right)=1$$

where the last equality follows from Proposition 13.2 and Proposition 13.5. So 17 is a quadratic residue modulo 19. Now suppose $x_0^2\equiv 17 \pmod{19}$. Then $x_0\not\equiv 0 \pmod{19}$, and hence $f'(x_0)\equiv 4x_0^3\cdot 2x_0=8x_0^4\not\equiv 0 \pmod{19}$. So by Hensel's Lemma, x_0 lifts to a solution to $x^2-17\equiv 0 \pmod{19^r}$ for any $r\geq 1$. Hence $f(x)\equiv 0 \pmod{19^r}$ has a solution for any $r\geq 1$.

Now we are done. □

We conclude with a baby example that uses techniques in Diophantine approximation. The main feature is the use of inequalities.

Proposition 18.6. *The equation* $x^4+x^3+x^2+x+1 = y^2$ *has integer solutions* $(-1,1),(0,1),(3,11)$ *and no others.*

Proof. Let $f(x) = 4x^4 + 4x^3 + 4x^2 + 4x + 4$. Note that any solution (x_0, y_0) of the original equation gives a solution $(x_0, 2y_0)$ to the equation $y^2 = f(x)$. So we only need to study the equation $y^2 = f(x)$.

Fist we have

$$f(x) = (2x^2 + x)^2 + 3(x + \frac{2}{3})^2 + \frac{8}{3} > (2x^2 + x)^2$$

and

$$f(x) = (2x^2 + x + 1)^2 - (x + 1)(x - 3).$$

Hence if $(x + 1)(x - 3) > 0$, or equivalently if $x > 3$ or $x < -1$, then $f(x)$ is between two consecutive squares and so itself cannot be a square! Thus any solution to $y^2 = f(x)$ must satisfy $-1 \leq x \leq 3$. There are then 5 possibilities for x. Checking them all gives the conclusion. □

19. Pell Equation $x^2 - 2y^2 = m$.

19.1. **The ring $\mathbb{Z}[\sqrt{2}]$.** Let $\sqrt{2}$ be a root of x^2-2 in $\mathbb{C}$ and let $\mathbb{Z}[\sqrt{2}] = \{a+b\sqrt{2} : a, b \in \mathbb{Z}\} \subset \mathbb{C}$.

The addition and multiplication formula for $\mathbb{Z}[\sqrt{2}]$ is as follows:

$$\text{Addition: } (a + b\sqrt{2}) + (c + d\sqrt{2}) = (a + c) + (b + d)\sqrt{2}.$$

$$\text{Multiplication: } (a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2}.$$

From this, we can check that $(\mathbb{Z}[\sqrt{2}], +, \cdot; 0, 1)$ is a commutative ring.

We define the **norm** in $\mathbb{Z}[\sqrt{2}]$ to be $N(a+b\sqrt{2}) = (a+b\sqrt{2})(a-b\sqrt{2}) = a^2 - 2b^2$. Note that unlike $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$, $N(\alpha)$ in $\mathbb{Z}[\sqrt{2}]$ may be negative. (So we will need a modification for the division algorithm in $\mathbb{Z}[\sqrt{2}]$. See Proposition 19.5.)

It is not easy to decide $\mathbb{Z}[\sqrt{2}]^\times$, the set of invertible elements of $\mathbb{Z}[\sqrt{2}]$. We will do this later. First let us prove a simple property for elements in $\mathbb{Z}[\sqrt{2}]^\times$.

Lemma 19.1. *We have $N(\alpha\beta) = N(\alpha)N(\beta)$ for any $\alpha, \beta \in \mathbb{Z}[\sqrt{2}]$.*

Proof. Write $\alpha = a + b\sqrt{2}$ and $\beta = c + d\sqrt{2}$. Then

$N(\alpha\beta) = N((ac+2bd)+(ad+bc)\sqrt{2}) = (ac+2bd)^2-2(ad+bc)^2 = (a^2-2b^2)(c^2-2d^2) = N(\alpha)N(\beta)$. □

Proposition 19.2. $\alpha \in \mathbb{Z}[\sqrt{2}]^\times \Leftrightarrow N(\alpha) = \pm 1$.

Proof. If $\alpha \in \mathbb{Z}[\sqrt{2}]^\times$, then $\alpha\beta = 1$ for some $\beta \in \mathbb{Z}[\sqrt{2}]$. Then $N(\alpha)N(\beta) = N(\alpha\beta) = 1$. So $N(\alpha) = \pm 1$.

Conversely if $N(\alpha) = \pm 1$, then $\alpha(a - b\sqrt{2}) = \pm 1$ (write $\alpha = a + b\sqrt{2}$). So either $a - b\sqrt{2}$ or $-a + b\sqrt{2}$ is the inverse of α. □

To illustrate the fact that $\mathbb{Z}[\sqrt{2}]$ is not as easy as previous cases ($\mathbb{Z}[i]$, $\mathbb{Z}[\sqrt{-2}]$), let us prove the following simple lemma.

Lemma 19.3. *We have*

$$\{\pm(1 + \sqrt{2})^n : n \in \mathbb{Z}\} = \{\pm 1, (\pm 1 + \sqrt{2})^n, -(\pm 1 + \sqrt{2})^n : n \text{ positive integer}\} \subset \mathbb{Z}[\sqrt{2}]^\times.$$

Proof. Since $(1+\sqrt{2})^n(-1+\sqrt{2})^n = ((1+\sqrt{2})(-1+\sqrt{2}))^n = 1^n = 1$, we have that $(-1+\sqrt{2})^{-n} = (1 + \sqrt{2})^n$ and so

$$\{(1 + \sqrt{2})^n : n \in \mathbb{Z}\} = \{1, (\pm 1 + \sqrt{2})^n : n \text{ positive integer}\} \subset \mathbb{Z}[\sqrt{2}]^\times.$$

Similarly $\{-(1 + \sqrt{2})^n : n \in \mathbb{Z}\} = \{-1, -(\pm 1 + \sqrt{2})^n : n \text{ positive integer}\} \subset \mathbb{Z}[\sqrt{2}]^\times$. □

Remark 19.4. *The deep reason for this, which we cannot prove in this class, is that $i, \sqrt{-2}$ are square roots of negative integers but $\sqrt{2}$ is a square root of a positive integer. In fact this is a general fact: If we consider the ring obtained from $\mathbb{Z}$ and a square root of a square-free integer D,*[8] *then its set of invertible elements is finite if $D < 0$ and is infinite if $D > 0$. In the latter case, this set of invertible elements equals $\{\pm\alpha^n : n \in \mathbb{Z}\}$ for some element α.*

Later on, we will use the following fact, which we do not prove:

$$\mathbb{Z}[\sqrt{2}]^\times = \{\pm(1 + \sqrt{2})^n : n \in \mathbb{Z}\} = \{\pm 1, (\pm 1 + \sqrt{2})^n, -(\pm 1 + \sqrt{2})^n : n \text{ positive integer}\}.$$

[8] This ring is $\mathbb{Z}[\frac{1+\sqrt{D}}{2}]$ if $D \equiv 1 \pmod 4$ and $\mathbb{Z}[\sqrt{D}]$ otherwise.

There are two ways to prove it: either use continued fractions or by geometry of numbers. The first proof is elementary, the second proof is more general (works for other more complicated situations in Algebraic Number Theory).

Now let us turn to the Division Algorithm and UFT for $\mathbb{Z}[\sqrt{2}]$.

Proposition 19.5 (Division Algorithm for $\mathbb{Z}[\sqrt{2}]$)**.** *We have the following division algorithm for $\mathbb{Z}[\sqrt{2}]$: For any $\alpha, \beta \in \mathbb{Z}[\sqrt{2}]$ with $\beta \neq 0$, there exist elements $\gamma, \delta \in \mathbb{Z}[\sqrt{2}]$ such that*

$$\alpha = \beta\gamma + \delta, \qquad \text{with } |N(\delta)| < |N(\beta)|.$$

Proof. Write $\alpha = a + b\sqrt{2}$ and $\beta = c + d\sqrt{2}$. Then

$$\frac{\alpha}{\beta} = \frac{a+b\sqrt{2}}{c+d\sqrt{2}} = \frac{(a+b\sqrt{2})(c-d\sqrt{2})}{(c+d\sqrt{2})(c-d\sqrt{2})} = \frac{ac-2bd}{c^2-2d^2} + \frac{bc-ad}{c^2-2d^2}\sqrt{2} =: s + t\sqrt{2}.$$

Let $x, y \in \mathbb{Z}$ such that $|x - s| \leq 1/2$ and $|y - t| \leq 1/2$. Set $\gamma = x + y\sqrt{2}$ and $\delta = \alpha - \beta\gamma$. Then

$$|N(\delta)| = |N(\beta)N((s-x)+(t-y)\sqrt{2}| = |N(\beta)|(|s-x|^2+2|t-y|^2) \leq |N(\beta)|(1/4+1/2) < |N(\beta)|.$$

□

Now since $|N(\alpha)| \geq 0$ for any $\alpha \in \mathbb{Z}[\sqrt{2}]$, this division algorithm indeed implies the Euclid's algorithm. So there are two equivalent definitions for prime elements in $\mathbb{Z}[\sqrt{2}]$, and $\mathbb{Z}[\sqrt{2}]$ satisfies the Unique Factorization Theorem.

Definition 19.6. *A nonzero non-invertible element α of $\mathbb{Z}[\sqrt{2}]$ is a prime element if any divisor of α is the multiple of 1 or α by an invertible element.*

Definition 19.7. *A nonzero non-invertible element α of $\mathbb{Z}[\sqrt{2}]$ is a prime element if the following condition holds: $\alpha|\beta\gamma$ implies $\alpha|\beta$ or $\alpha|\gamma$.*

Theorem 19.8 (UFT for $\mathbb{Z}[\sqrt{2}]$)**.** *Every nonzero element α of $\mathbb{Z}[\sqrt{2}]$ can be written as the product of prime elements with an invertible element, unique up to reordering. More precisely for any $\alpha \in \mathbb{Z}[\sqrt{2}]$, we have*

1. *$\alpha = \mathfrak{p}_1 \cdots \mathfrak{p}_n$ for some prime elements $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ of $\mathbb{Z}[\sqrt{2}]$;*
2. *If $\alpha = \mathfrak{p}_1 \cdots \mathfrak{p}_n = \mathfrak{p}'_1 \cdots \mathfrak{p}'_m$ for prime elements $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ and $\mathfrak{p}'_1, \ldots, \mathfrak{p}'_m$, then $n = m$ and up to reordering $\mathfrak{p}_i \in \mathfrak{p}'_i\mathbb{Z}[\sqrt{2}]^\times$ for each $i \in \{1, \ldots, n\}$.*

Next let us see some propositions regarding the prime factorizations.

Proposition 19.9. *If the norm of $\alpha \in \mathbb{Z}[\sqrt{2}]$ is $\pm p$ for a prime number p, then α is a prime element.*

Proof. If $\alpha = \beta\gamma$, then $\pm p = N(\alpha) = N(\beta)N(\gamma)$. So either $N(\beta) = \pm 1$ or $N(\gamma) = \pm 1$. So by Proposition 19.2, either $\beta \in \mathbb{Z}[\sqrt{2}]^\times$ or $\gamma \in \mathbb{Z}[\sqrt{2}]^\times$. Hence we are done. □

Example 19.10. *Write down a prime factorization of 7 in $\mathbb{Z}[\sqrt{2}]$. One possible way is $7 = (3+\sqrt{2})(3-\sqrt{2})$. By the previous exercise, both $3+\sqrt{2}$ and $3-\sqrt{2}$ are prime elements because their norms are 7 which is a prime number.*

19.2. **Relation between $\mathbb{Z}[\sqrt{2}]$ and the Diophantine equations $x^2-2y^2=m$.** Let us first see some examples.

Example 19.11. *Find all integer solutions to $x^2-2y^2=1$.*

Proof. We have that $x^2-2y^2=1$ is equivalent to $N(x+y\sqrt{2})=1$. Hence we are looking for elements $\alpha:=x+y\sqrt{2}\in\mathbb{Z}[\sqrt{2}]$ such that $N(\alpha)=1$.

By Proposition 19.2, such α's are elements of $\mathbb{Z}[\sqrt{2}]^\times$. Now let us check the norms of the elements in $\mathbb{Z}[\sqrt{2}]^\times$. Since $N(1+\sqrt{2})=-1$, we have that

$$N(\pm(1+\sqrt{2})^n)=N(\pm1)N((1+\sqrt{2})^n)=1(-1)^n=(-1)^n.$$

So

$$\{\alpha\in\mathbb{Z}[\sqrt{2}]:N(\alpha)=1\}=\{\pm(1+\sqrt{2})^{2m}:m\in\mathbb{Z}\}=\{\pm1,(\pm1+\sqrt{2})^{2m},-(\pm1+\sqrt{2})^{2m}:m\text{ positive integer}\}.$$

So the integer solutions to $x^2-2y^2=1$ are

$$(\pm1,0),(\sum_{k=0}^{m}\binom{2m}{2k}2^k,\pm\sum_{k=1}^{m}\binom{2m}{2k-1}2^{k-1}),(-\sum_{k=0}^{m}\binom{2m}{2k}2^k,\pm\sum_{k=1}^{m}\binom{2m}{2k-1}2^{k-1}).$$

□

Example 19.12. *Find all integer solutions to $x^2-2y^2=-1$.*

Proof. Similarly to the previous example, the integer solutions to $x^2-2y^2=-1$ are those (x,y) corresponding to the elements of

$$\{\alpha\in\mathbb{Z}[\sqrt{2}]:N(\alpha)=-1\}=\{\pm(1+\sqrt{2})^{2m-1}:m\in\mathbb{Z}\}=\{(\pm1+\sqrt{2})^{2m-1},-(\pm1+\sqrt{2})^{2m-1}:m\text{ positive integer}\}$$

by $\alpha=x+y\sqrt{2}$. Now expand. □

Example 19.13. *Find all integer solutions to $x^2-2y^2=2$.*

Proof. The integer solutions to $x^2-2y^2=2$ are those (x,y) corresponding to the elements of $\{\alpha\in\mathbb{Z}[\sqrt{2}]:N(\alpha)=2\}$ by $\alpha=x+y\sqrt{2}$.

We compute $\{\alpha\in\mathbb{Z}[\sqrt{2}]:N(\alpha)=2\}$. First $N(\sqrt{2})=-2$. So

$$\{\pm\sqrt{2}(1+\sqrt{2})^{2m-1}:m\in\mathbb{Z}\}\subset\{\alpha\in\mathbb{Z}[\sqrt{2}]:N(\alpha)=2\}.$$

We will prove that this inclusion is actually an equality. Then it suffices to expand.

For any $\alpha=a+b\sqrt{2}\in\mathbb{Z}[\sqrt{2}]$ such that $N(\alpha)=2$, denote by $\beta=a-b\sqrt{2}$. Then $\alpha\beta=2=(\sqrt{2})^2$. So $\sqrt{2}|\alpha\beta$. But $\sqrt{2}$ is a prime element because $N(\sqrt{2})=-2$ (by Proposition 19.9), so either $\sqrt{2}|\alpha$ or $\sqrt{2}|\beta$.

But $N(\alpha)=N(\beta)=2$, so α and β are prime elements by Proposition 19.9. So $\alpha\in\sqrt{2}\mathbb{Z}[\sqrt{2}]^\times$ or $\beta\in\sqrt{2}\mathbb{Z}[\sqrt{2}]^\times$. In either case we have $\alpha\in\sqrt{2}\mathbb{Z}[\sqrt{2}]^\times$ because $\alpha=\beta=-(\sqrt{2})^2$. So $\alpha=\sqrt{2}u$ for some $u\in\mathbb{Z}[\sqrt{2}]^\times$. But $N(\alpha)=2$ and $N(\sqrt{2})=-2$, so $N(u)=-1$. Therefore $\alpha\in\{\sqrt{2}u:N(u)=-1\}$. Hence

$$\{\alpha\in\mathbb{Z}[\sqrt{2}]:N(\alpha)=2\}\subset\{\pm\sqrt{2}(1+\sqrt{2})^{2m-1}:m\in\mathbb{Z}\}.$$

This is what we want. □

Example 19.14. *Find all integer solutions to $x^2-2y^2=-2$.*

Proof. Similar argument as above. We have

$$\{\alpha \in \mathbb{Z}[\sqrt{2}] : N(\alpha) = -2\} = \{\sqrt{2}u : N(u) = 1\} = \{\pm\sqrt{2}(1+\sqrt{2})^{2m} : m \in \mathbb{Z}\}.$$

□

Example 19.15. *Find all integer solutions to* $x^2 - 2y^2 = \pm 3$.

Proof. Suppose (x_0, y_0) is an integer solution. First $x_0 \neq 0$ and $y_0 \neq 0$. The square of any integer modulo 3 is 0 or 1, so $x_0^2 - 2y_0^2 \equiv 0 \pmod 3$ if and only if $x_0 \equiv y_0 \equiv 0 \pmod 3$. Write $x_0 = 3x_1$ and $y_0 = 3y_1$ for $x_1, y_1 \in \mathbb{Z}$, then $3(x_1^2 - 2y_1^2) = \pm 1$. But this cannot happen. □

Corollary 19.16. 3 *is a prime element of* $\mathbb{Z}[\sqrt{2}]$.

Proof. If $3 = \alpha\beta$, then $N(\alpha)N(\beta) = N(3) = 9$. We want to prove that either α or β is invertible, so it suffices to prove that we cannot have $N(\alpha) = \pm 3$. But then it follows from Example 19.15. □

Example 19.17. *Find all integer solutions to* $x^2 - 2y^2 = 7$.

Proof. Write $(x + \sqrt{2}y)(x - \sqrt{2}y) = (3 + \sqrt{2})(3 - \sqrt{2})$. The right hand side is a prime decomposition.

Let $\alpha = x + \sqrt{2}y$. Then $N(\alpha) = 7$ and so α is a prime element. So $\alpha = u(3 \pm \sqrt{2})$ with $u \in \mathbb{Z}[\sqrt{2}]^\times$ (similar argument as before). Comparing the norms of both sides we have $N(u) = 1$. So the elements $\alpha \in \mathbb{Z}[\sqrt{2}]$ we are looking for are precisely

$$(1+\sqrt{2})^{2m}(3 \pm \sqrt{2}),\ -(1+\sqrt{2})^{2m}(3 \pm \sqrt{2}) \quad \text{with } m \in \mathbb{Z}.$$

□

Example 19.18. *Find all integer solutions to* $x^2 - 2y^2 = 49$.

Proof. We are looking for elements $\alpha \in \mathbb{Z}[\sqrt{2}]$ such that $N(\alpha) = 49$.

We want to look for a prime decomposition of α. First note that α is not a prime element. This is because $\alpha(a - b\sqrt{2}) = 49 = (3 + \sqrt{2})^2(3 - \sqrt{2})^2$ (where we write $\alpha = a + b\sqrt{2}$) is the product of four prime elements.

Now for any decomposition $\alpha = \beta\gamma$, we have $N(\beta)N(\gamma) = N(\alpha) = 49 = 7^2$. Assume $N(\beta) \neq \pm 1$, then we are in one of the two cases: $N(\beta) = N(\gamma) = \pm 7$ or $N(\beta) = \pm 49$ and $N(\gamma) = \pm 1$. There is a 1-to-1 correspondence between elements of norm 49 and elements of norm -49, by sending $\alpha \mapsto (1 + \sqrt{2})\alpha$.

So it suffices to investigate the case $N(\beta) = N(\gamma) = \pm 7$. Similar to the previous exercise, we can show that

$$\{\beta \in \mathbb{Z}[\sqrt{2}] : N(\beta) = 7\} = \{(1+\sqrt{2})^{2m}(3 \pm \sqrt{2}),\ -(1+\sqrt{2})^{2m}(3 \pm \sqrt{2}) \quad \text{with } m \in \mathbb{Z}\},$$

$$\{\beta \in \mathbb{Z}[\sqrt{2}] : N(\beta) = -7\} = \{(1+\sqrt{2})^{2m-1}(3 \pm \sqrt{2}),\ -(1+\sqrt{2})^{2m-1}(3 \pm \sqrt{2}) \quad \text{with } m \in \mathbb{Z}\}.$$

Then the elements α we are looking for are precisely the products of any two elements in one of the set above (but they must come from the same set). □

Proposition 19.19. *Now let* p *be any prime number. The following statements are equivalent:*

(1) $x^2 - 2y^2 = p$ *has integer solutions;*

(2) $x^2 - 2y^2 = -p$ *has integer solutions;*

(3) p is not a prime element in $\mathbb{Z}[\sqrt{2}]$.

Moreover $x^2 - 2y^2 = p$ has either no integer solutions or infinitely many integer solutions (same for $x^2 - 2y^2 = -p$).

Proof. First note that $x + \sqrt{2}y \in \mathbb{Z}[\sqrt{2}]^\times \Leftrightarrow x - \sqrt{2}y \in \mathbb{Z}[\sqrt{2}]^\times \Leftrightarrow -x + \sqrt{2}y \in \mathbb{Z}[\sqrt{2}]^\times$. So (1) $\Rightarrow$ (3) and (2) $\Rightarrow$ (3).

On the other hand, it is not hard to prove that (1) $\Leftrightarrow$ (2): If $N(a + \sqrt{2}b) = p$, then $N((a + 2b) + \sqrt{2}(a + b)) = N((a + \sqrt{2}b)(1 + \sqrt{2})) = -p$ and vice versa.

Now we prove that (3) implies (1) or (2). Suppose p is not a prime element in $\mathbb{Z}[\sqrt{2}]$, then $p = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[\sqrt{2}]$ not invertible. Then $p^2 = N(p) = N(\alpha)N(\beta)$. Now $N(\alpha) = N(\beta) = \pm p$ because $N(\alpha) \neq \pm 1$ and $N(\beta) \neq \pm 1$. So if we write $\alpha = a + b\sqrt{2}$, then (a, b) is an integer solution to $x^2 - 2y^2 = \pm p$. Hence (3) implies (1) or (2).

Now if (a, b) is an integer solution to $x^2 - 2y^2 = p$, then any element $(a + b\sqrt{2})u$ with $N(u) = 1$ gives an integer solution to $x^2 - 2y^2 = p$. Since any $u = \pm(1 + \sqrt{2})^n$ with n even has norm 1, so there are infinitely many integer solutions to $x^2 - 2y^2 = p$. Exactly the same argument for $x^2 - 2y^2 = -p$. $\square$

20. $x^3 + y^3 = z^3$ WITH $3 \nmid xyz$

The goal of this section is to find all the integer solutions to the following Fermat equation

$$x^3 + y^3 = z^3.$$

There are some obvious solutions: $(0, k, k)$, $(k, 0, k)$ and $(k, -k, 0)$ for all $k \in \mathbb{Z}$. We call these *trivial solutions*.

Theorem 20.1. *There are no non-trivial integer solutions to $x^3 + y^3 = z^3$ with $3 \nmid xyz$.*

The claim is also true when we remove the assumption $3 \nmid xyz$. Its proof is harder and we will consider it later.

Suppose (a, b, c) is an integer solution. One can factor $a^3 + b^3$ as

$$a^3 + b^3 = (a + b)(a^2 - ab + b^2).$$

We wish to furthermore decompose $a^2 - ab + b^2 = b^2\left((\frac{a}{b})^2 - \frac{a}{b} + 1\right)$. To do this let $\zeta \in \mathbb{C}$ be a solution to $x^2 - x + 1 = 0$. There are two choices for ζ. In this lecture we take

$$\zeta = \frac{1 + \sqrt{-3}}{2}.$$

Then the other solution to $x^2 - x + 1 = 0$ is $\overline{\zeta} = 1 - \zeta$.

20.1. **The ring $\mathbb{Z}[\zeta]$.** We now consider the set $\mathbb{Z}[\zeta] = \{a + b\zeta : a, b \in \mathbb{Z}\} \subset \mathbb{C}$. Recall that ζ satisfies $\zeta^2 = \zeta - 1$. Then the addition and multiplication on $\mathbb{Z}[\zeta]$ are given by the formulae

$$\begin{aligned}(a + b\zeta) + (c + d\zeta) &= (a + c) + (b + d)\zeta,\\ (a + b\zeta)(c + d\zeta) &= ac + (ad + bc)\zeta + bd\zeta^2\\ &= ac + (ad + bc)\zeta + bd(\zeta - 1)\\ &= (ac - bd) + (ad + bc + bd)\zeta.\end{aligned}$$

Then $(\mathbb{Z}[\zeta], +, \cdot; 0, 1)$ is a commutative ring.

Next let us define the norm in $\mathbb{Z}[\zeta]$. We define

$$N(a+b\zeta) = (a+b\zeta)\overline{(a+b\zeta)} = (a+\frac{b}{2}+\frac{b}{2}\sqrt{-3})(a+\frac{b}{2}-\frac{b}{2}\sqrt{-3}) = (a+\frac{b}{2})^2+3(\frac{b}{2})^2 = a^2+ab+b^2.$$

Then $N(a + b\zeta) \in \mathbb{Z}$, $N(a + b\zeta) \geq 0$ and $N(a + b\zeta) = 0 \Leftrightarrow a = b = 0$. Again we have

$$N(\alpha)N(\beta) = (\alpha\overline{\alpha})(\beta\overline{\beta}) = (\alpha\beta)\overline{(\alpha\beta)} = N(\alpha\beta).$$

Lemma 20.2. *For any $\alpha \in \mathbb{Z}[\zeta]$, we have $\alpha \in \mathbb{Z}[\zeta]^\times \Leftrightarrow N(\alpha) = 1$.*

Proof. If $N(\alpha) = 1$, then $\alpha\overline{\alpha} = 1$ by the definition of norms. So $\alpha \in \mathbb{Z}[\zeta]^\times$.

Conversely if $\alpha \in \mathbb{Z}[\zeta]^\times$, then $\alpha\beta = 1$ for some $\beta \in \mathbb{Z}[\zeta]$. Taking norms of both sides we have $N(\alpha)N(\beta) = 1$. But $N(\alpha) \geq 1$ and $N(\beta) \geq 1$, so $N(\alpha) = N(\beta) = 1$. □

Proposition 20.3. $\mathbb{Z}[\zeta]^\times = \{1, -1, \zeta, -\zeta, \overline{\zeta} = 1 - \zeta, -\overline{\zeta} = -1 + \zeta\}$.

Proof. We have $\zeta(1 - \zeta) = 1$. So the right hand side is contained in $\mathbb{Z}[\zeta]^\times$.

Conversely for any $\alpha = a+b\zeta \in \mathbb{Z}[\zeta]^\times$, we have $N(\alpha) = 1$ by Lemma 20.2. So $(a+\frac{b}{2})^2+\frac{3}{4}b^2 = 1$. But then we have $b = 0, \pm 1$.

If $b = 0$, then $a = \pm 1$. So $\alpha = \pm 1$.

If $b = 1$, then $a = 0$ or -1. So $\alpha = \zeta$ or $-1 + \zeta$.
If $b = -1$, then $a = 0$ or 1. So $\alpha = -\zeta$ or $1 - \zeta$. □

Next we turn to division algorithm.

Proposition 20.4 (Division Algorithm for $\mathbb{Z}[\zeta]$)**.** *For any $\alpha, \beta \in \mathbb{Z}[\zeta]$ with $\beta \neq 0$, there exist $\gamma, \delta \in \mathbb{Z}[\zeta]$ such that*

$$\alpha = \beta\gamma + \delta, \quad \text{with } N(\delta) < N(\beta).$$

Proof. Write $\alpha = a + b\zeta$ and $\beta = c + d\zeta$. Then

$$\begin{aligned}
\frac{a+b\zeta}{c+d\zeta} &= \frac{a + \frac{b}{2} + \frac{b\sqrt{-3}}{2}}{c + \frac{d}{2} + \frac{d\sqrt{-3}}{2}} \frac{c + \frac{d}{2} - \frac{d\sqrt{-3}}{2}}{c + \frac{d}{2} - \frac{d\sqrt{-3}}{2}} \\
&= \frac{ac + bd + \frac{bc+ad}{2} + \frac{bc-ad}{2}\sqrt{-3}}{c^2 + cd + d^2} \\
&= \frac{ac + bd + ad}{c^2 + cd + d^2} + \frac{bc - ad}{c^2 + cd + d^2}\zeta \\
&=: s + t\zeta.
\end{aligned}$$

Here s, t are rational numbers. Let x, y be the closest integer to s, t respectively. Then $|x - s| \leq \frac{1}{2}$ and $|y - t| \leq \frac{1}{2}$. Set $\gamma = x + y\zeta$ and $\delta = \alpha - \beta\gamma$. Then $\delta = \beta((s - x) + (t - y)\zeta)$ and hence

$$\begin{aligned}
N(\delta) &= N(\beta)N((s - x) + (t - y)\zeta) \\
&= N(\beta)\left((s - x)^2 + (s - x)(t - y) + (t - y)^2\right) \\
&\leq N(\beta)(|s - x|^2 + |s - x||t - y| + |t - y|^2) \\
&\leq N(\beta)(\frac{1}{4} + \frac{1}{4} + \frac{1}{4}) < N(\beta).
\end{aligned}$$

□

With the Division Algorithm in hand, we obtain the Euclid's Algorithm and the notion of gcd (defined by $\alpha\mathbb{Z}[\zeta] + \beta\mathbb{Z}[\zeta] = \gcd(\alpha, \beta)\mathbb{Z}[\zeta]$ and is well-defined up to $\mathbb{Z}[\zeta]^\times$). So $\mathbb{Z}[\zeta]$ is a principal ideal domain (PID). Moreover we have two equivalent definitions of prime elements of $\mathbb{Z}[\zeta]$, and the Unique Factorization Theorem.

We end this subsection with the following proposition.

Proposition 20.5. *If $\alpha \in \mathbb{Z}[\zeta]$ satisfies that $N(\alpha)$ is a prime number, then α is a prime element of $\mathbb{Z}[\zeta]$.*

The proof is the same as for $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$.

20.2. **Application to $x^3 + y^3 = z^3$.** Now we are ready to prove Theorem 20.1. Recall that ζ is a solution to $x^2 - x + 1 = 0$ and (a, b, c) is an integer solution to $x^3 + y^3 = z^3$. We assume that $3 \nmid abc$. Then (a, b, c) is clearly non-trivial.

If p divides two of a, b, c, then it also divides the third. Hence $(a/p, b/p, c/p)$ is another integer solution to $x^3 + y^3 = z^3$. Hence by dividing out the common factor, we may assume that a, b, c are pairwise coprime.

The solutions to $x^2 - x + 1 = 0$ are, by computation, ζ and $1 - \zeta$. So

$$x^2 - x + 1 = (x - \zeta)(x - (1 - \zeta)).$$

So we can complete our factorization of a^3+b^3 in $\mathbb{Z}[\zeta]$:

$$a^3+b^3=(a+b)(a^2-ab+b^2)=(a+b)(a-b\zeta)(a-b(1-\zeta)).$$

Lemma 20.6. *$a+b, a-b\zeta$ and $a-b(1-\zeta)$ are pairwise coprime in $\mathbb{Z}[\zeta]$, namely any element in $\mathbb{Z}[\zeta]$ dividing two of them is in $\mathbb{Z}[\zeta]^\times$.*

Proof. We start with the first two. Note that

$$\gcd(a+b,a-b\zeta)=\gcd(a+b,(a+b)-(a-b\zeta))=\gcd(a+b,b(1+\zeta)).$$

By assumption a and b are coprime in $\mathbb{Z}$, so $a+b$ and b are coprime in $\mathbb{Z}$. So $1=(a+b)m+bn$ for some $m,n\in\mathbb{Z}$. But then $1\in(a+b)\mathbb{Z}[\zeta]+b\mathbb{Z}[\zeta]$, and hence $\mathbb{Z}[\zeta]=(a+b)\mathbb{Z}[\zeta]+b\mathbb{Z}[\zeta]$. So $a+b$ and b are also coprime in $\mathbb{Z}[\zeta]$. So by the UFT for $\mathbb{Z}[\zeta]$, we have $\gcd(a+b,b(1+\zeta))=\gcd(a+b,1+\zeta)$. Thus

$$\gcd(a+b,a-b\zeta)=\gcd(a+b,1+\zeta).$$

On the other hand we have $N(1+\zeta)=3$. So $1+\zeta$ is a prime element of $\mathbb{Z}[\zeta]$ by Proposition 20.5. So $\gcd(a+b,a-b\zeta)=1$ or $1+\zeta$. We only need to exclude the latter case.

Suppose $1+\zeta|a+b$, then $N(1+\zeta)|N(a+b)$.[9] So $3|(a+b)^2$. But then $3|a+b$ as 3 is a prime number. But then $c^3=a^3+b^3\equiv(a+b)^3 \pmod 3$ implies that $3|c$. This contradicts our assumption that $3\nmid abc$. Hence we proved that $a+b$ and $a-b\zeta$ are coprime in $\mathbb{Z}[\zeta]$.

The proof for $a+b$ and $a-b(1-\zeta)$ is similar: we have $\gcd(a+b,a-b(1-\zeta))=\gcd(a+b,b(2-\zeta))=\gcd(a+b,2-\zeta)$ and $N(2-\zeta)=3$.

Now we prove that $a-b\zeta$ and $a-b(1-\zeta)$ are coprime in $\mathbb{Z}[\zeta]$. We have

$$\gcd(a-b\zeta,a-b(1-\zeta))=\gcd(a-b\zeta,b(1-2\zeta)).$$

But $\gcd(a-b\zeta,b)=\gcd(a-b\zeta+b\zeta,b)=\gcd(a,b)=1$. So

$$\gcd(a-b\zeta,a-b(1-\zeta))=\gcd(a-b\zeta,1-2\zeta).$$

We have $N(1-2\zeta)=3$. So $1-2\zeta$ is a prime element of $\mathbb{Z}[\zeta]$. So it suffices to prove $1-2\zeta\nmid a-b\zeta$. Suppose otherwise, then we have $N(1-2\zeta)|N(a-b\zeta)$, namely

$$3|a^2-ab+b^2=(a+b)^3-3ab.$$

So again $3|a+b$ and consequently c. Contradiction! □

Now let us study the equality

$$c^3=a^3+b^3=(a+b)(a-b\zeta)(a-b(1-\zeta)).$$

The left hand side is a cube. The right hand is the product of pairwise coprime elements of $\mathbb{Z}[\zeta]$. Hence by UFT for $\mathbb{Z}[\zeta]$, there exist elements $\alpha_1,\alpha_2,\alpha_3\in\mathbb{Z}[\zeta]$ and $u_1,u_2,u_3\in\mathbb{Z}[\zeta]^\times=\{\pm1,\pm\zeta,\pm(1-\zeta)\}$ such that

$$\begin{aligned}a+b&=u_1\alpha_1^3,\\ a-b\zeta&=u_2\alpha_2^3,\\ a-b(1-\zeta)&=u_3\alpha_3^3.\end{aligned}$$

[9]For $a+b=(1+\zeta)\alpha$, we have $a+b=\overline{(1+\zeta)}\overline{\alpha}$ by taking the complex conjugation. Then $N(a+b)=(a+b)^2=(1+\zeta)\alpha\overline{(1+\zeta)}\overline{\alpha}=N(1+\zeta)N(\alpha)$.

Note that for any $\alpha = s + t\zeta \in \mathbb{Z}[\zeta]$, we have

$$\alpha^3 \equiv s^3 + t^3\zeta^3 = s^3 - t^3 \equiv s - t \pmod 3.$$

Here modulo 3 is done in $\mathbb{Z}[\zeta]$ instead of in $\mathbb{Z}$! [10]

Let us write $\alpha_2 = s + t\zeta$. Then $a - b\zeta = u_2\alpha_2^3$ implies

$$a - b\zeta \equiv u_2(s - t) \pmod 3.$$

Since $s, t \in \mathbb{Z}$, we must have $u_2 = \pm(1 - \zeta)$. Up to replacing α_2 by $-\alpha_2$ we may assume $u_2 = 1 - \zeta$.

Now by using $\zeta^2 = \zeta - 1$, we have

$$\begin{aligned}
a - b\zeta &= (1 - \zeta)(s + t\zeta)^3 \\
&= (1 - \zeta)(s^3 + 3s^2t\zeta + 3st^2\zeta^2 + t^3\zeta^3) \\
&= (1 - \zeta)(s^3 + 3s^2t\zeta + 3st^2(\zeta - 1) - t^3) \\
&= (1 - \zeta)\left(s^3 - 3st^2 - t^3 + (3s^2t + 3st^2)\zeta\right) \\
&= s^3 - 3st^2 - t^3 + (3s^2t + 3st^2 - s^3 + 3st^2 + t^3)\zeta + (-3s^2t - 3st^2)\zeta^2 \\
&= s^3 - 3st^2 - t^3 + (3s^2t + 3st^2 - s^3 + 3st^2 + t^3)\zeta + (-3s^2t - 3st^2)(\zeta - 1) \\
&= (s^3 + 3s^2t - t^3) + (-s^3 + 3st^2 + t^3)\zeta.
\end{aligned}$$

Comparing both sides, we have $a = s^3 + 3s^2t - t^3 \equiv s^3 - t^3 \pmod 3$ and $b = -s^3 + 3st^2 + t^3 \equiv -s^3 + t^3 \pmod 3$. So $a \equiv -b \pmod 3$. But then

$$c^3 = a^3 + b^3 \equiv (-b)^3 + b^3 = 0 \pmod 3.$$

So $3|c^3$ and hence $3|c$. Contradiction to $3 \nmid abc$. This proves Theorem 20.1.

Remark 20.7. *After we proved $u_2 = 1 - \zeta$, the first natural way to continue the argument is as follows: we have then $a - b\zeta \equiv (1 - \zeta)(s - t) \pmod 3$, and hence*

$$a - (s - t) \equiv (b - (s - t))\zeta \pmod 3.$$

The left hand side does not have ζ and the right hand side is a multiple of ζ. So both sides are 0. Hence

$$a \equiv b \equiv s - t \pmod 3.$$

As a, b and $s - t$ are integers, this congruence can be seen in $\mathbb{Z}$ or in $\mathbb{Z}[\zeta]$ (same effect). Unfortunately this does NOT give any contradiction. So we must go through the complicated computation as we did.

[10] As a set we have that $\mathbb{Z}[\zeta]/3\mathbb{Z}[\zeta] = \{a + b\zeta : a, b = 0, 1, 2\}$ has 9 elements.

21. Solving $x^3+y^3=z^3$ in $\mathbb{Z}[\zeta]$

The goal is to prove the following improvement of Theorem 20.1.

Theorem 21.1. *There are no non-trivial solutions to $x^3+y^3=z^3$ in $\mathbb{Z}[\zeta]$, namely: if (α,β,γ) is a solution in $\mathbb{Z}[\zeta]$, then $\alpha\beta\gamma=0$.*

If (α,β,γ) is a solution to $x^3+y^3=z^3$ in $\mathbb{Z}[\zeta]$ and $0\neq\delta\in\mathbb{Z}[\zeta]$ divides two out of three, then δ also divides the third one. Then $(\alpha/\delta,\beta/\delta,\gamma/\delta)$ gives another solution to $x^3+y^3=z^3$ in $\mathbb{Z}[\zeta]$. Hence if (α,β,γ) is a non-trivial solution, then we may assume that α,β,γ are pairwise coprime in $\mathbb{Z}[\zeta]$.

21.1. **Local analysis: The modular arithmetic $\mathbb{Z}[\zeta]/(\sqrt{-3})$.** Let us consider the element $\sqrt{-3}\in\mathbb{Z}[\zeta]$. We have $N(\sqrt{-3})=3$, and hence $\sqrt{-3}$ is a prime element of $\mathbb{Z}[\zeta]$ by Proposition 20.5. Then $3=-(\sqrt{-3})^2$ is a prime factorization of 3 in $\mathbb{Z}[\zeta]$.

Suppose $\alpha\in\mathbb{Z}[\zeta]$ satisfies $N(\alpha)=3$, then $-(\sqrt{-3})^2=3=\alpha\overline{\alpha}$. So $\sqrt{-3}|\alpha$ or $\overline{\alpha}$. But $\overline{\sqrt{-3}}=-\sqrt{-3}$, so $\sqrt{-3}|\alpha$. So up to an element in $\mathbb{Z}[\zeta]^\times$, α is $\sqrt{-3}$ or 3. But $N(3)=9\neq 3$. So $\alpha\in\sqrt{-3}\mathbb{Z}[\zeta]^\times$.[11]

We want to study the modular arithmetic $\mathbb{Z}[\zeta]/\sqrt{-3}\mathbb{Z}[\zeta]$, which we denote by $\mathbb{Z}[\zeta]/(\sqrt{-3})$ for simplicity. First we point out that the ring structure on $\mathbb{Z}[\zeta]$ induces a ring structure on $\mathbb{Z}[\zeta]/(\sqrt{-3})$ (as we did for modular arithmetic in $\mathbb{Z}$ or $F[x]$).

Proposition 21.2. *As rings, $(\mathbb{Z}[\zeta]/(\sqrt{-3}),+,\cdot;\overline{0},\overline{1})$ and $(\mathbb{Z}/3\mathbb{Z},+,\cdot;[0],[1])$ are isomorphic. In particular, $(\mathbb{Z}[\zeta]/(\sqrt{-3}),+,\cdot;\overline{0},\overline{1})$ is a field.*

Proof. We give two proofs for this proposition. The first proof is explicit, while the second proof is abstract but more general.

Method 1 We define the following map

$$\phi\colon \mathbb{Z}/3\mathbb{Z}\to\mathbb{Z}[\zeta]/(\sqrt{-3}),\ [a]\mapsto\overline{a}.$$

We prove that ϕ is well-defined: If $a=a'+3k$ for some $k\in\mathbb{Z}$, then $a=a'+\sqrt{-3}(1-2\zeta)k$ and hence $\overline{a}=\overline{a'}$. We prove that ϕ is a ring homomorphism: $\phi([0])=\overline{0}$, $\phi([1])=\overline{1}$, $\phi([a]+[b])=\phi([a])+\phi([b])=\overline{a+b}$ and $\phi([a][b])=\phi([a])\phi([b])=\overline{ab}$.

We prove that ϕ is injective: Suppose $\phi([a])=\phi([b])$, then $\overline{a}=\overline{b}$. So $a=b+\sqrt{-3}\alpha$ for some $\alpha\in\mathbb{Z}[\zeta]$. But then $\sqrt{-3}\alpha=a-b\in\mathbb{Z}$. Taking the norms of both sides, we get $3|3N(\alpha)=N(\sqrt{3}\alpha)=N(a-b)=(a-b)^2$. Since 3 is a prime number, we have $3|a-b$. Thus $[a]=[b]$.

We prove that ϕ is surjective: By a simple computation we have $\zeta=2+\sqrt{-3}\zeta$. Hence $\overline{\zeta}=\overline{2}=\phi([2])$. So for any $\alpha=a+b\zeta\in\mathbb{Z}[\zeta]$, we have $\overline{\alpha}=\phi([a+2b])$.

Now ϕ^{-1} is the desired isomorphism.

Method 2 We first prove that $\mathbb{Z}[\zeta]/(\sqrt{-3})$ is finite. That is, there are finitely many congruence classes modulo $\sqrt{-3}$. Note that if two elements of $\mathbb{Z}[\zeta]$ are congruent modulo 3, then they are congruent modulo $\sqrt{-3}$ since $\sqrt{-3}|3$ in $\mathbb{Z}[\zeta]$. Hence the number of congruence classes modulo $\sqrt{3}$ is at most the number of congruence classes modulo 3, which is 9. This proves the finiteness of $\mathbb{Z}[\zeta]/(\sqrt{-3})$. But we have more: since $\sqrt{-3}$ and 0 are congruent modulo $\sqrt{-3}$ but not modulo 3, we see that $\#(\mathbb{Z}[\zeta]/(\sqrt{-3}))<9$.

[11]For example, $1+\zeta=\sqrt{-3}(1-\zeta)$ and $1-2\zeta=\sqrt{-3}(-1)$.

Next we shall use some abstract facts, which we prove later. Since $\sqrt{-3}$ is a prime element, we have that $\mathbb{Z}[\zeta]/(\sqrt{-3})$ is an integral domain, namely the product of any two non-zero elements is still non-zero. But $\mathbb{Z}[\zeta]/(\sqrt{-3})$ is finite, and hence it is a field. These follow from the two claims below.

Since 3 is congruent to 0 modulo $\sqrt{-3}$, we see that the field $\mathbb{Z}[\zeta]/(\sqrt{-3})$ is of characteristic 3. But $\#(\mathbb{Z}[\zeta]/(\sqrt{-3})) < 9 = 3^2$. So $\#(\mathbb{Z}[\zeta]/(\sqrt{-3})) = 3$ by Theorem 10.5. Now $\mathbb{Z}[\zeta]/(\sqrt{-3})$ is a field with 3 elements. So it must be $\mathbb{F}_3$.

Claim 1. *Let R be an integral domain, and let α be a prime element, namely $\alpha|\beta\gamma \Rightarrow \alpha|\beta$ or $\alpha|\gamma$. Then $R/\alpha R$ is again an integral domain.*

Claim 2. *Let R be an integral domain. If R has finitely many elements, then R is a field.*

Proof of Claim 1. Suppose $\overline{\beta}\overline{\gamma} = \overline{0_R}$. Then $\beta\gamma \in \alpha R$, or equivalently $\alpha|\beta\gamma$. So $\alpha|\beta$ or $\alpha|\gamma$, or equivalently $\overline{\beta} = \overline{0_R}$ or $\overline{\gamma} = \overline{0_R}$. Thus $R/\alpha R$ is an integral domain by definition. □

Proof of Claim 2. Take any $0 \neq \alpha \in R$. We wish to prove that $\alpha\beta = 1_R$ for some $\beta \in R$.

Let us define a map $\psi\colon R \to R,\ \beta \mapsto \alpha\beta$. This map is injective: If $\psi(\beta) = \psi(\beta')$, then $\alpha\beta = \alpha\beta'$. Hence $\alpha(\beta - \beta') = 0_R$. But $\alpha \neq 0_R$ and R is an integral domain. So $\beta - \beta' = 0_R$. So $\beta = \beta'$.

So ψ is also surjective because R is finite. In particular $1_R = \psi(\beta)$ for some $\beta \in R$. This β is what we desire. □

□

Proposition 21.2 implies that any $\alpha \in \mathbb{Z}[\zeta]$ is congruent to 0 or 1 or -1 modulo $\sqrt{-3}$.

Proposition 21.3. *Let $e \in \{1, -1\}$. If $\alpha \in \mathbb{Z}[\zeta]$ is congruent to e modulo $\sqrt{-3}$, then α^3 is congruent to e modulo 9.*

Proof. Write $\alpha = e + \beta\sqrt{-3}$ for $\beta \in \mathbb{Z}[\zeta]$. Then we have

$$\begin{aligned}\alpha^3 &= e^3 + 3e^2\beta\sqrt{-3} + 3e(\beta\sqrt{-3})^2 + (\beta\sqrt{-3})^3 \\ &= e + 9e\beta^2 + 3\sqrt{-3}(\beta - \beta^3) \\ &\equiv e + 3\sqrt{-3}\beta(\beta - 1)(\beta + 1) \pmod 9.\end{aligned}$$

By Proposition 21.2, β is congruent to 0 or 1 or -1 modulo $\sqrt{-3}$. So $\sqrt{-3}|\beta(\beta-1)(\beta+1)$, and hence $9|3\sqrt{-3}\beta(\beta-1)(\beta+1)$. Thus $\alpha^3 \equiv e \pmod 9$. □

Corollary 21.4. *Suppose $\alpha, \beta, \gamma \in \mathbb{Z}[\zeta]$ pairwise coprime and $\epsilon \in \mathbb{Z}[\zeta]^\times$ satisfy $\alpha^3 + \beta^3 + \epsilon\gamma^3 = 0$. Then $\sqrt{-3}$ divides exactly one of α, β, γ.*

Proof. Since α, β, γ are pairwise coprime and $\sqrt{-3} \notin \mathbb{Z}[\zeta]^\times$, we know that $\sqrt{-3}$ divides at most one of α, β, γ.

If $\sqrt{-3}$ does not divide any one of them, then $\alpha^3, \beta^3, \gamma^3$ are congruent to 1 or -1 modulo 9 by Proposition 21.3. But then $\alpha^3 + \beta^3 + \epsilon\gamma^3$ cannot be congruent to 0 modulo 9. Contradiction. □

Corollary 21.5. *Suppose $\alpha, \beta \in \mathbb{Z}[\zeta]$ neither of which is divisible by $\sqrt{-3}$. Suppose $\epsilon \in \mathbb{Z}[\zeta]^\times = \{\pm 1, \pm\zeta, \pm(1-\zeta)\}$ with $\alpha^3 + \epsilon\beta^3 \equiv 0 \pmod 3$. Then $\epsilon = \pm 1$ and moreover $\alpha^3 + \epsilon\beta^3 \equiv 0 \pmod 9$.*

Proof. By Proposition 21.3, we have $\alpha^3, \beta^3 \equiv \pm 1 \pmod 9$. Now the only elements in $\{\pm 1, \pm\zeta, \pm(1-\zeta)\}$ for which $\pm 1 \pm \epsilon$ can possibly be congruent to 0 modulo 3 are ± 1. □

21.2. **Global analysis: A descent argument.** In order to prove Theorem 21.1, we need to prove an even stronger form of it.

Theorem 21.6. *There do not exist (pairwise) coprime elements $\alpha, \beta, \gamma \in \mathbb{Z}[\zeta]$ and an element $\epsilon \in \mathbb{Z}[\zeta]^\times$ such that*

$$\alpha^3 + \beta^3 + \epsilon\gamma^3 = 0. \tag{21.1}$$

We will prove this theorem by a descent argument: we show that if one has a quadruple $(\alpha, \beta, \gamma, \epsilon)$ satisfying the desired conditions, then there exists another quadruple $(\alpha', \beta', \gamma', \epsilon')$ also satisfying the desired conditions but with $N(\alpha'\beta'\gamma') < N(\alpha\beta\gamma)$. This process can be repeated forever, but $N(\alpha\beta\gamma)$ is always a positive integer. This cannot happen. Thus we get a contradiction.

First of all, $\sqrt{-3}$ divides exactly one of α, β, γ by Corollary 21.4. If $\sqrt{-3}$ divides α or β, then $\epsilon = \pm 1 = \epsilon^3$ by Corollary 21.5. So in this case we can absorb ϵ into γ^3, namely replace γ by $\epsilon\gamma$. Then α, β, γ are still (pairwise) coprime and we have $\alpha^3 + \beta^3 + \gamma^3 = 0$ in this case. We can then rename α, β, γ so that $\sqrt{-3}|\gamma$.

In summary, we may assume $\sqrt{-3}|\gamma$ and so $\sqrt{-3} \nmid \alpha, \beta$. Moreover, $\alpha^3 + \beta^3 \equiv 0 \pmod 9$ by Corollary 21.5. So $(\sqrt{-3})^4 = 9|\gamma^3$. Since $\sqrt{-3}$ is a prime element in $\mathbb{Z}[\zeta]$, we have $(\sqrt{-3})^2|\gamma$ by UFT for $\mathbb{Z}[\zeta]$.

Similar to last time, we factor $\alpha^3 + \beta^3$ as

$$-\epsilon\gamma^3 = \alpha^3 + \beta^3 = (\alpha + \beta)(\alpha - \beta\zeta)(\alpha - \beta(1 - \zeta)). \tag{21.2}$$

Lemma 21.7. *If $\varpi \notin \mathbb{Z}[\zeta]^\times$ divides any two of $\alpha+\beta$, $\alpha-\beta\zeta$ or $\alpha-\beta(1-\zeta)$, then $\varpi \in \sqrt{-3}\mathbb{Z}[\zeta]^\times$.*

Proof. Suppose ϖ divides the first two. Then ϖ divides $\alpha(1 + \zeta) = (\alpha + \beta)\zeta + (\alpha - \beta\zeta)$ and $\beta(1 + \zeta) = (\alpha + \beta) - (\alpha - \beta\zeta)$. Since α, β are coprime, we get $\varpi|1 + \zeta$. Since $N(1 + \zeta) = 3$, we have that $1 + \zeta$ is a prime element and is in $\sqrt{-3}\mathbb{Z}[\zeta]^\times$. Hence $\varpi \in \sqrt{-3}\mathbb{Z}[\zeta]^\times$.

Suppose ϖ divides the first and the third, then the same argument shows $\varpi|2-\zeta$. Once again $\varpi \in \sqrt{-3}\mathbb{Z}[\zeta]^\times$ since $N(2 - \zeta) = 3$.

Suppose ϖ divides the last two, then $\varpi|1 - 2\zeta = -\sqrt{-3}$ and so $\varpi \in \sqrt{-3}\mathbb{Z}[\zeta]^\times$. □

We introduce the following notation: for any element $\delta \in \mathbb{Z}[\zeta]$, we denote by $\mu_{\sqrt{-3}}(\delta)$ to be the largest non-negative integer k such that $(\sqrt{-3})^k|\delta$.

Recall that $\sqrt{-3} \nmid \beta$. So the difference of any two of $\alpha + \beta$, $\alpha - \beta\zeta$ and $\alpha - \beta(1 - \zeta)$ is not by $\sqrt{-3}^2$. Write

$$\begin{aligned} a &= \mu_{\sqrt{-3}}(\alpha + \beta) \\ b &= \mu_{\sqrt{-3}}(\alpha - \beta\zeta) \\ c &= \mu_{\sqrt{-3}}(\alpha - \beta(1 - \zeta)). \end{aligned}$$

(The proof of) Lemma 21.7 implies

$$\alpha + \beta \equiv \alpha - \beta\zeta \equiv \alpha - \beta(1 - \zeta) \pmod{\sqrt{-3}}.$$

Thus either $a = b = c = 0$ or $a, b, c \geq 1$. But $\sqrt{-3}^6 | -\epsilon\gamma^3 = (\alpha + \beta)(\alpha + \beta)(\alpha - \beta(1 - \zeta))$, so $a + b + c \geq 6$. Thus $a, b, c \geq 1$.

On the other hand at most one of a, b, c is ≥ 2 since the difference of any two of $\alpha+\beta$, $\alpha-\beta\zeta$ and $\alpha - \beta(1 - \zeta)$ is not divisible by $\sqrt{-3}^2$.

Thus as multi-sets $\{a,b,c\} = \{1,1,3r-2\}$ where $r = \mu_{\sqrt{-3}}(\gamma) \geq 2$.

Dividing both sides of (21.2) by $\sqrt{-3}^{3r}$, we get

$$-\epsilon\left(\frac{\gamma}{\sqrt{-3}^r}\right)^3 = \frac{\alpha+\beta}{\sqrt{-3}^a}\frac{\alpha-\beta\zeta}{\sqrt{-3}^b}\frac{\alpha-\beta(1-\zeta)}{\sqrt{-3}^c}.$$

Lemma 21.7 implies that the three terms on the right hand side are pairwise coprime. Hence by UFT for $\mathbb{Z}[\zeta]$, there exist $\epsilon_1,\epsilon_2,\epsilon_3 \in \mathbb{Z}[\zeta]^\times$ and pairwise coprime elements $\lambda_1,\lambda_2,\lambda_3 \in \mathbb{Z}[\zeta]$ such that

$$\begin{aligned}\alpha+\beta &= \epsilon_1\sqrt{-3}^a\lambda_1^3,\\ \alpha-\beta\zeta &= \epsilon_2\sqrt{-3}^b\lambda_2^3,\\ \alpha-\beta(1-\zeta) &= \epsilon_3\sqrt{-3}^c\lambda_3^3.\end{aligned}$$

Since $\zeta, 1-\zeta \in \mathbb{Z}[\zeta]^\times$, we multiply the second equation by $-\zeta$ and the last by $-(1-\zeta)$ and get

$$\begin{aligned}\alpha+\beta &= \epsilon_1\sqrt{-3}^a\lambda_1^3,\\ -\zeta\alpha-\beta(1-\zeta) &= \epsilon_4\sqrt{-3}^b\lambda_2^3,\\ -(1-\zeta)\alpha-\beta\zeta &= \epsilon_5\sqrt{-3}^c\lambda_3^3,\end{aligned}$$

where $\epsilon_4 = -\epsilon_2\zeta$, $\epsilon_5 = -\epsilon_3(1-\zeta) \in \mathbb{Z}[\zeta]^\times$. Summing up these three new equations gives

$$0 = \epsilon_1\sqrt{-3}^a\lambda_1^3 + \epsilon_4\sqrt{-3}^b\lambda_2^3 + \epsilon_5\sqrt{-3}^c\lambda_3^3.$$

Since this equation is symmetric, we may assume without loss of generality that $a=1$, $b=1$, $c=3r-2$. Dividing by $\epsilon_1\sqrt{-3}$ gives

$$0 = \lambda_1^3 + \epsilon_6\lambda_2^3 + \epsilon_7(\sqrt{-3}^{r-1}\lambda_3)^3 = 0$$

for some $\epsilon_6,\epsilon_7 \in \mathbb{Z}[\zeta]^\times$. Since $r \geq 2$, we have $\lambda_1^3 + \epsilon_6\lambda_2^3 \equiv 0 \pmod 3$. So by Corollary 21.5, $\epsilon_6 = \pm 1 = \epsilon_6^3$. Set $(\alpha',\beta',\gamma',\epsilon') = (\lambda_1, \epsilon_6\lambda_2, \sqrt{-3}^{r-1}\lambda_3, \epsilon_7)$. Then $(\alpha',\beta',\gamma',\epsilon')$ is a quadruple satisfying (21.1).

Since

$$\lambda_1^3\lambda_2^3(\sqrt{-3}^{r-1}\lambda_3)^3 = \frac{u_1(\alpha^3+\beta^3)}{\sqrt{-3}^3} = \frac{u_2\gamma^3}{\sqrt{-3}^3}$$

for some $u_1,u_2 \in \mathbb{Z}[\zeta]^\times$, we have

$$N((\alpha'\beta'\gamma')^3) = N(\gamma^3/\sqrt{-3}^3) < N(\gamma^3).$$

So $N(\alpha'\beta'\gamma') < N(\gamma) \leq N(\alpha\beta\gamma)$.

22. Quadratic Fields

22.1. Algebraic Numbers and Algebraic Integers.

Definition 22.1. *A complex number ξ is called an **algebraic number** if it is a root of some non-zero $f \in \mathbb{Q}[x]$.*

For example any rational number r is algebraic since we can take $f(x) = x - r$. There are many other examples, like $\sqrt{D}$ for any $D \in \mathbb{Q}$ as we can take $f(x) = x^2 - D$.

Any complex number that is not algebraic is said to be **transcendental**. The best known examples of transcendental numbers are π and e.

Theorem 22.2. *Let ξ be an algebraic number. Then there exists a unique irreducible monic polynomial $f \in \mathbb{Q}[x]$ such that $f(\xi) = 0$. Moreover f satisfies the following property: For any $g \in \mathbb{Q}[x]$ such that $g(\xi) = 0$, we have $f|g$.*

Proof. By definition of algebraic numbers, ξ is a root of some non-zero polynomial in $\mathbb{Q}[x]$. Let f be such a polynomial of minimal degree, say d. Such an f exists because the degree of a non-zero polynomial is non-negative. We may assume that f is monic by dividing f by the leading coefficient.

We prove that f is irreducible. Suppose $f = f_1 f_2$ with $f_1, f_2 \in \mathbb{Q}[x]$. It suffices to prove that either $\deg(f_1) = 0$ or $\deg(f_2) = 0$. Observe that $f_1 \neq 0$ and $f_2 \neq 0$ since $f \neq 0$. Now $0 = f(\xi) = f_1(\xi)f_2(\xi)$. So $f_1(\xi) = 0$ or $f_2(\xi) = 0$. But then $\deg(f_1) \geq \deg(f)$ or $\deg(f_2) \geq \deg(f)$ by the minimality of $\deg(f)$. However $\deg(f) = \deg(f_1) + \deg(f_2) \geq \max(\deg(f_1), \deg(f_2))$. So either $\deg(f_2) = 0$ or $\deg(f_1) = 0$.

Now for any $g \in \mathbb{Q}[x]$ such that $g(\xi) = 0$, we apply the Division Algorithm and get $g = fh + r$ for some $h, r \in \mathbb{Q}[x]$ with $\deg(r) < \deg(f)$. Then $0 = g(\xi) = f(\xi)h(\xi) + r(\xi) = r(\xi)$. By the minimality of $\deg(f)$, we must then have $r = 0$. Hence $f|g$. This proves the "Moreover" part of the theorem.

We turn to the uniqueness of f. Let $f^* \in \mathbb{Q}[x]$ be another irreducible monic polynomial with $f^*(\xi) = 0$. Then by the previous paragraph we have $f|f^*$. So $f^* = fh$ for some $h \in \mathbb{Q}[x]$. But f^* is irreducible and f is non-constant, so $h \in \mathbb{Q}^*$. By comparing leading coefficients we get $h = 1$. So $f^* = f$. □

Definition 22.3. *The **minimal polynomial** of an algebraic number ξ is the polynomial $f(x) \in \mathbb{Q}[x]$ described in Theorem 22.2. The **degree** of ξ, denoted by $\deg(\xi)$, is defined to be $\deg(f)$.*

Definition 22.4. *Let ξ be an algebraic number and let $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$ be its minimal polynomial. Suppose the roots of f in $\mathbb{C}$ are $\xi_1 = \xi, \ldots, \xi_n$. Then the **norm** of ξ, denoted by $N(\xi)$, is defined to be $\xi_1 \cdots \xi_n = (-1)^n a_0 \in \mathbb{Q}$. The **trace** of ξ, denoted by $\mathrm{Tr}(\xi)$, is defined to be $\xi_1 + \cdots + \xi_n = -a_{n-1} \in \mathbb{Q}$.*

We have seen norms in $\mathbb{Z}[i], \mathbb{Z}[\sqrt{-2}], \mathbb{Z}[\sqrt{2}]$ and $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$. All these norms are particular cases of the norm defined here if the element is not an integer. There is another way to define norms and traces: we first fix a field and then talk about norms and traces in that field. These two definitions are slightly different. For example $N(2) = 2$ for Definition 22.4, but $N(2) = 2^2 = 4$ if we use the norm in $\mathbb{Z}[i]$ defined previously.

Definition 22.5. *An algebraic number ξ is an **algebraic integer** if it is a root of some monic $f \in \mathbb{Z}[x]$.*

In this definition, it is important to request f to be monic: Otherwise let $f \in \mathbb{Q}[x]$ non-zero be such that $f(\xi) = 0$ and let $f^* = cf$ where $c \in \mathbb{Z}$ is a multiple of the denominator of any coefficient of f (e.g. c can be taken to be the lcm of the denominators of the coefficients). Then ξ is a root of $f^* \in \mathbb{Z}[x]$. But we certainly do not want any algebraic number to be an algebraic integer.

The following lemma justifies the name "integer" in the sense that it coincides with the common sense in the simplest case.

Lemma 22.6. *A rational number is an algebraic integer if and only if it is in $\mathbb{Z}$.*

Proof. It is clear that any integer is an algebraic integer.

On the other hand if $\frac{a}{b} \in \mathbb{Q}$ is an algebraic integer, then we may assume $\gcd(a, b) = 1$. By definition of algebraic integers

$$\left(\frac{a}{b}\right)^n + c_{n-1}\left(\frac{a}{b}\right)^{n-1} + \ldots + c_0 = 0$$

for some $c_0, \ldots, c_{n-1} \in \mathbb{Z}$. Multiplying both sides by b^n we get

$$a^n + c_{n-1}ba^{n-1} + \ldots + c_0 b^n = 0.$$

Thus $b|a^n$. But $\gcd(a, b) = 1$. Hence $b = \pm 1$. Thus $\frac{a}{b} \in \mathbb{Z}$. □

Proposition 22.7. *The minimal polynomial of an algebraic integer is in $\mathbb{Z}[x]$.*

Proof. Suppose ξ is an algebraic integer. Then $g(\xi) = 0$ for some monic polynomial $g \in \mathbb{Z}[x]$.

Let $f \in \mathbb{Q}[x]$ be the minimal polynomial of ξ. Then by Theorem 22.2, we have $g = fh$ for some $h \in \mathbb{Q}[x]$. Since both f and g are monic, h is also monic.

By Gauß's Lemma (Lemma 22.8), we have $f \in \mathbb{Z}[x]$ (and $h \in \mathbb{Z}[x]$). Hence we are done. □

Lemma 22.8 (Gauß's Lemma)**.** *Let $f \in \mathbb{Z}[x]$ be monic. Suppose $f = f_1 f_2$ for $f_1, f_2 \in \mathbb{Q}[x]$ monic, then $f_1, f_2 \in \mathbb{Z}[x]$.*

Proof. We introduce the following notion: A polynomial $g \in \mathbb{Z}[x]$ is said to be **primitive** if the only positive integer dividing all its coefficients is 1. We make use of the following Claim.

Claim 3. *The product of two primitive polynomials is primitive.*

Let c_1 be the smallest positive integer such that $c_1 f_1(x) \in \mathbb{Z}[x]$. Then $c_1 f_1(x)$ is a primitive polynomial: If p divides all coefficients of $c_1 f_1(x)$, then $p|c_1$ since c_1 is the leading coefficient, and hence $(c_1/p)f_1(x) \in \mathbb{Z}[x]$, contradiction to the minimality of c_1.

Similarly let c_2 be the smallest positive integer such that $c_2 f_2(x) \in \mathbb{Z}[x]$. Then $c_2 f_2(x)$ is a primitive polynomial.

Then $c_1 c_2 f(x) = (c_1 f_1(x))(c_2 f_2(x))$ is primitive by Claim 3. But $f \in \mathbb{Z}[x]$, so $c_1 c_2 = 1$. Hence $c_1 = c_2 = 1$. So it suffices to prove Claim 3.

Proof of Claim 3. Let $a_n x^n + \ldots + a_0$ and $b_m x^m + \ldots + b_0$ be primitive polynomials and let denote their product by $c_{n+m}x^{n+m} + \ldots + c_0$. Suppose that the product is not primitive. Then there exists a prime number p such that $p|c_k$ for each $k \in \{0, \ldots, n+m\}$.

Let a_i be the first coefficient of $a_n x^n + \ldots + a_0$ such that $p \nmid a_i$. Such an i exists since $a_n x^n + \ldots + a_0$ is primitive. Let b_j be the first coefficient of $b_m x^m + \ldots + b_0$ such that $p \nmid b_j$.

Such a j exists since $b_m x^m + \ldots + b_0$ is primitive. Then the coefficient of x^{i+j} in the product polynomial is

$$c_{i+j} = \sum_{0\le k\le n,\ 0\le i+j-k\le m} a_k b_{i+j-k}.$$

In this sum, any term with $k > i$ is a multiple of p by looking at a_k, any term with $i+j-k > j$ (i.e. $i > k$) is a multiple of p by looking at b_k. Hence $c_{i+j} \equiv a_i b_j \pmod p$. Hence $p | a_i b_j$. This contradicts $p \nmid a_i$ and $p \nmid b_j$. □

□

We close this subsection by the following important proposition.

Proposition 22.9. *Let α and β be algebraic numbers (resp. algebraic integers). Then $\alpha+\beta$ and $\alpha\beta$ are also algebraic numbers (resp. algebraic integers).*

Proof. (Sketch) Let f be the minimal polynomial of α and let g be the minimal polynomial of β. Suppose the roots of f in $\mathbb{C}$ are $\alpha_1 = \alpha, \alpha_2, \ldots, \alpha_n$ and the roots of g in $\mathbb{C}$ are $\beta_1 = \beta, \beta_2, \ldots, \beta_m$. Then

$$f(x) = \prod_{i=1}^{n}(x-\alpha_i) \quad \text{and} \quad g(x) = \prod_{j=1}^{m}(x-\beta_j).$$

Let

$$h_1(x) = \prod_{1\le i\le n,\ 1\le j\le m} (x-(\alpha_i+\beta_j)) \quad \text{and} \quad h_2(x) = \prod_{1\le i\le n,\ 1\le j\le m} (x-\alpha_i\beta_j)$$

Then $h_1, h_2 \in \mathbb{Q}[x]$. (Not very easy. Expand for j, then for i.)[12]

If α and β are algebraic integers, then $f, g \in \mathbb{Z}[x]$ by Proposition 22.7. Then $h_1, h_2 \in \mathbb{Z}[x]$. □

Corollary 22.10. *The set of all algebraic numbers forms of a field. The set of all algebraic integers forms a commutative ring.*

Proof. If α, β are algebraic numbers, then so are $\alpha+\beta$ and $\alpha\beta$ by Proposition 22.9.

If α is a root of $a_n x^n + \ldots + a_0$, then $-\alpha$ is a root of $(-1)^n a_n x^n + \ldots + a_0$. If furthermore $\alpha \neq 0$, then α^{-1} is a root of $a_n + a_{n-1}x + \ldots + a_0 x^n$.

Hence the set of all algebraic numbers is closed under addition, multiplication, negation and inverse. It certainly contains 0 and 1. Hence it is a field.

As for the set of all algebraic integers, it is closed under addition, multiplication, negation but not necessarily inverse. It contains 0 and 1. Hence it is a commutative ring. □

22.2. **Number Fields.** For any algebraic number ξ, define

$$\mathbb{Q}(\xi) = \left\{ \frac{f(\xi)}{h(\xi)} : f, h \in \mathbb{Q}[x],\ h(\xi) \neq 0 \right\}.$$

Here the inverse is taken in $\mathbb{C}$. On $\mathbb{C}$, we have the addition and multiplication. Now

$$\frac{f_1(\xi)}{h_1(\xi)} + \frac{f_2(\xi)}{h_2(\xi)} = \frac{f_1(\xi)h_2(\xi) + f_2(\xi)h_1(\xi)}{h_1(\xi)h_2(\xi)} = \frac{(f_1h_2 + f_2h_1)(\xi)}{(h_1h_2)(\xi)} \in \mathbb{Q}(\xi),$$

[12]Philosophically, h_1 and h_2 are stable under all possible "conjugations" because we have added all possible conjugates of $\alpha+\beta$ and $\alpha\beta$ respectively, so they are in $\mathbb{Q}[x]$. This argument can be made precise by Galois theory. Otherwise it is also possible to expand these two polynomials and prove directly that they are in $\mathbb{Q}[x]$.

$$\frac{f_1(\xi)}{h_1(\xi)}\frac{f_2(\xi)}{h_2(\xi)} = \frac{(f_1f_2)(\xi)}{(h_1h_2)(\xi)} \in \mathbb{Q}(\xi).$$

So $(\mathbb{Q}(\xi), +, \cdot; 0, 1)$ is a commutative ring. It is easy to check that this ring is actually a field.

Lemma 22.11. *Let $n = \deg(\xi)$. Then every element of $\mathbb{Q}(\xi)$ can be written uniquely as*

$$a_0 + a_1\xi + \ldots + a_{n-1}\xi^{n-1}$$

where $a_i \in \mathbb{Q}$ for all $i \in \{0, \ldots, n-1\}$.

Proof. (existence) Let $g \in \mathbb{Q}[x]$ be the minimal polynomial of ξ. In particular g is irreducible and $\deg(g) = n$.

For any $\frac{f(\xi)}{h(\xi)} \in \mathbb{Q}(\xi)$, we have $h(\xi) \neq 0$. But $g(\xi) = 0$, so $g \nmid h$. But g is irreducible, so $\gcd(g, h) = 1$. Hence $g(x)G(x) + h(x)H(x) = 1$ for some $G, H \in \mathbb{Q}[x]$. Take $x = \xi$ in this equation, then we have $h(\xi)H(\xi) = 1$. Hence $\frac{f(\xi)}{h(\xi)} = f(\xi)H(\xi) = (fH)(\xi)$.

Now apply the Division Algorithm in $\mathbb{Q}[x]$ to fH and g, we get $fH = gq + r$ for some $q, r \in \mathbb{Q}[x]$ with $\deg(r) < \deg(g) = n$. Then $r(\xi) = (fH)(\xi)$. Hence $r(\xi)$ is the expression which we desire.

(uniqueness) If $a_0+a_1\xi+\ldots+a_{n-1}\xi^{n-1} = a'_0+a'_1\xi+\ldots+a'_{n-1}\xi^{n-1}$ with $a_0, \ldots, a_{n-1}, a'_0, \ldots, a'_{n-1} \in \mathbb{Q}$, then ξ is a root of the polynomial

$$g^*(x) = (a_{n-1} - a'_{n-1})x^{n-1} + \ldots + (a_0 - a'_0) \in \mathbb{Q}[x].$$

Now $\deg(g^*) \leq n-1$ and $\deg(\xi) = n$, so $g^* = 0$. Hence $a_i = a'_i$ for all $i \in \{0, \ldots, n-1\}$. □

Another way to look at $\mathbb{Q}(\xi)$ is to use the Modular Arithmetic in $\mathbb{Q}[x]$. Note that we have used this point of view for the classification of finite fields. See the proof of Theorem 11.6.

Theorem 22.12. *Let g be the minimal polynomial of ξ. Then there exists a ring (or field) isomorphism*

$$(\mathbb{Q}(\xi), +, \cdot; 0, 1) \simeq (\mathbb{Q}[x]/(g), +, \cdot; \overline{0}, \overline{1}).$$

Proof. Define a map

$$\varphi\colon \mathbb{Q}[x]/(g) \to \mathbb{Q}(\xi),\ \overline{f} \mapsto f(\xi).$$

We prove that φ is well-defined: If $\overline{f}_1 = \overline{f}_2$, then $f_1 - f_2 = gh$ for some $h \in \mathbb{Q}[x]$. Then $f_1(\xi) - f_2(\xi) = g(\xi)h(\xi) = 0$. Hence $f_1(\xi) = f_2(\xi)$.

We prove that φ is a ring homomorphism: $\varphi(\overline{0}) = 0$ and $\varphi(\overline{1}) = 1$. Also $\varphi(\overline{f}_1 + \overline{f}_2) = \varphi(\overline{f_1 + f_2}) = (f_1 + f_2)(\xi) = f_1(\xi) + f_2(\xi) = \varphi(\overline{f}_1) + \varphi(\overline{f}_2)$. Similarly $\varphi(\overline{f}_1\overline{f}_2) = \varphi(\overline{f}_1)\varphi(\overline{f}_2)$.

We prove that φ is injective: If $\varphi(\overline{f}_1) = \varphi(\overline{f}_2)$, i.e. $f_1(\xi) = f_2(\xi)$, then $(f_1 - f_2)(\xi) = 0$. Hence $g | f_1 - f_2$ by the definition of minimal polynomials. Hence $\overline{f}_1 = \overline{f}_2$.

Finally φ is surjective by Lemma 22.11. □

22.3. **Ring of Integers.** For any algebraic number ξ, define the **ring of integers** of the field $\mathbb{Q}(\xi)$ to be the set of algebraic integers contained in $\mathbb{Q}(\xi)$. It is a ring by Proposition 22.9. This ring is usually denoted by $\mathcal{O}_{\mathbb{Q}(\xi)}$. It is clearly an integral domain, i.e. the product of any non-zero elements is still non-zero.

We use $\mathcal{O}^\times_{\mathbb{Q}(\xi)}$ to denote the set of invertible elements of $\mathcal{O}_{\mathbb{Q}(\xi)}$. Then $(\mathcal{O}^\times_{\mathbb{Q}(\xi)}, \cdot; 1)$ is an abelian group.

22.4. **Quadratic Fields.** Now we turn to quadratic fields. Let D be a square-free integer, and let $\sqrt{D}$ be a root of $x^2 - D$. We wish to understand $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$.

Theorem 22.13. $\mathcal{O}_{\mathbb{Q}(\sqrt{D})} = \begin{cases} \mathbb{Z}[\sqrt{D}] & \text{if } D \equiv 2,3 \pmod 4 \\ \mathbb{Z}[\frac{1+\sqrt{D}}{2}] & \text{if } D \equiv 1 \pmod 4 \end{cases}$.

Proof. We start with $\subset$. Let $\alpha \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$. Then $\alpha = a + b\sqrt{D}$ for some $a, b \in \mathbb{Q}$ by Lemma 22.11. Define $\alpha' = a - b\sqrt{D}$. Then $\alpha + \alpha' = 2a$ and $\alpha\alpha' = a^2 - Db^2$. Let $f(x) = x^2 - 2ax + (a^2 - Db^2)$. Then α, α' are the roots of f.

If $\alpha \in \mathbb{Q}$, then $\alpha \in \mathbb{Z}$ by Lemma 22.6. If $\alpha \notin \mathbb{Q}$, then $b \neq 0$. Hence f is irreducible in $\mathbb{Q}[x]$. So f is the minimal polynomial of α. But α is an algebraic integer, so $f \in \mathbb{Z}[x]$ by Proposition 22.7. So $2a \in \mathbb{Z}$ and $a^2 - Db^2 \in \mathbb{Z}$.

If $a \in \mathbb{Z}$, then $Db^2 \in \mathbb{Z}$. But D is square-free, so $b \in \mathbb{Z}$. Hence $\alpha \in \mathbb{Z}[\sqrt{D}]$ in this case.

If $a \in \frac{1}{2}\mathbb{Z} \setminus \mathbb{Z}$, then the denominator of a^2 is 4. So the denominator of Db^2 is also 4. Since D is square-free, we then have $b \in \frac{1}{2}\mathbb{Z} \setminus \mathbb{Z}$. Hence $\alpha = a + b\sqrt{D} = (a - b) + 2b\frac{1+\sqrt{D}}{2}$. Now $a - b \in \mathbb{Z}$ and $2b \in \mathbb{Z}$ since $a, b \in \frac{1}{2}\mathbb{Z} \setminus \mathbb{Z}$. Hence $\alpha \in \mathbb{Z}[\frac{1+\sqrt{D}}{2}]$. However, in this case we can rewrite

$$a^2 - Db^2 = \frac{(2a)^2 - D(2b)^2}{4} \in \mathbb{Z}.$$

So $(2a)^2 - D(2b)^2 \equiv 0 \pmod 4$. But $2a$ and $2b$ are odd integers, so we get $D \equiv 1 \pmod 4$.

To sum it up, we have the desired "$\subset$".

Conversely, $\sqrt{D}$ is a root of $x^2 - D \in \mathbb{Z}[x]$. So the "$\supset$" holds when $D \equiv 2,3 \pmod 4$. If $D \equiv 1 \pmod 4$, then $\frac{1+\sqrt{D}}{2}$ is a root of $x^2 - x + \frac{1-D}{4} \in \mathbb{Z}[x]$. So the "$\supset$" also holds when $D \equiv 1 \pmod 4$.

So we are done. □

Theorem 22.14. *For $D < -3$, we have $\mathcal{O}^\times_{\mathbb{Q}(\sqrt{D})} = \{\pm 1\}$.*

Proof. The inclusion $\supset$ is clear. We prove $\subset$. As before, we can prove that

$$\mathcal{O}^\times_{\mathbb{Q}(\sqrt{D})} = \{\alpha \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})} : N(\alpha) = \pm 1\}.$$

If $D \equiv 2,3 \pmod 4$, then we let $\alpha = a + b\sqrt{D} \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ with $a, b \in \mathbb{Z}$. We then have $N(\alpha) = a^2 - Db^2 \geq 0$ since $D < 0$. If $N(\alpha) = 1$, then $a^2 - Db^2 = 1$. Since $D < -3$ and so $-D > 3$, we must have $b = 0$ and $a = \pm 1$. So we are done in this case.

If $D \equiv 1 \pmod 4$, then we let $\alpha = a + b\frac{1+\sqrt{D}}{2} \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ with $a, b \in \mathbb{Z}$. We then have $N(\alpha) = (a + \frac{b}{2})^2 - D(\frac{b}{2})^2 \geq 0$ since $D < 0$. If $N(\alpha) = 1$, then $(a + \frac{b}{2})^2 - D(\frac{b}{2})^2 = 1$. Since $D < -3$ and so $-D > 3$, we must have $b = 0$ and $a = \pm 1$. hence we are done in this case. □

Remark 22.15. *When $D > 0$, we have the following result: There exists some $\alpha \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ such that*

$$\mathcal{O}^\times_{\mathbb{Q}(\sqrt{D})} = \{\pm \alpha^n : n \in \mathbb{Z}\}.$$

Theorem 22.16. *For $D = -1, -2, -3, -7, -11, 2, 3, 5$, the ring $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ has the Division Algorithm. Namely for any $\alpha, \beta \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ and $\beta \neq 0$, there exist $\gamma, \delta \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ such that*

$$\alpha = \beta\gamma + \delta, \quad \text{with } |N(\delta)| < |N(\beta)|.$$

The proof is similar as before. Note that $-7, -11, 5 \equiv 1 \pmod 4$. This is why they are not "too large" for the Division Algorithm to hold (recall that $\mathcal{O}_{\mathbb{Q}(\sqrt{-5})}$ cannot have the Division Algorithm as we have shown in §16.3).

Most of the time, we do not really need the Division Algorithm, but only that $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is PID. Here is a big theorem regarding this perspective.

Theorem 22.17. *If $D < 0$, then $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is PID if and only if $D = -1$, -2, -3, -7, -11, -19, -43, -67, -163.*

On the other hand, for $0 < D < 100$, $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is PID if and only if

$$\begin{aligned} D =& 2, 3, 5, 6, 7, 11, 13, 14, 17, 19, 21, 22, 23, 29, 31, 33, 37, 38, 41, 43, 46, 47, \\ & 53, 57, 59, 61, 62, 67, 69, 71, 73, 77, 83, 86, 89, 93, 94, 97. \end{aligned}$$

When an integral domain R is not a PID, then prime elements and irreducible elements are not the same. This is somewhat annoying. In order to (partly) fix this problem, we introduce the following definition of ideals and prime ideals.

Definition 22.18. *(1) An **ideal** I of R is a subset of R closed under addition such that $aI \subset I$ for any $a \in R$.*

*(2) An ideal I of R is called a **prime ideal** if the following property holds: For any $a, b \in R$, $ab \in I \Rightarrow a \in I$ or $b \in I$.*

For example, in all previous cases of PID ($R = \mathbb{Z}$, $F[x]$, $\mathbb{Z}[i]$, $\mathbb{Z}[\sqrt{-2}]$, $\mathbb{Z}[\sqrt{2}]$, $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$), what we actually cared for was αR for some element $\alpha \in R$. This is an ideal of R, which is usually denoted by (α) in the theory of rings. Then $\beta \in (\alpha) \Leftrightarrow \beta = \alpha\gamma$ for some $\gamma \in R \Leftrightarrow \alpha | \beta$. So the definition of prime ideal is a generalization of irreducible elements.[13] Moreover, $(\alpha) = R \Leftrightarrow \alpha \in R^\times$.

For elements $\alpha_1, \ldots, \alpha_n \in R$, we use $(\alpha_1, \ldots, \alpha_n)$ to denote the ideal $\alpha_1 R + \ldots + \alpha_n R$. If R is PID, then $(\alpha_1, \ldots, \alpha_n) = (\gamma)$ for some $\gamma \in R$.

We already know that $\mathcal{O}_{\mathbb{Q}(\sqrt{-5})}$ is not a PID. In fact we can prove that $(2, 1 + \sqrt{-5})$ cannot be expressed as (γ) for any $\gamma \in \mathcal{O}_{\mathbb{Q}(\sqrt{-5})}$. Suppose it could, then $2 \in (\gamma)$ and $1 + \sqrt{-5} \in (\gamma)$. Hence $2 = \gamma\alpha$ and $1 + \sqrt{-5} = \gamma\beta$. Taking the norms, we get

$$4 = N(\gamma)N(\alpha) \text{ and } 6 = N(\gamma)N(\beta).$$

So $N(\gamma) | \gcd(4, 6) = 2$. Hence $N(\gamma) = 1$ or 2.

If $N(\gamma) = 2$, then $a^2 + 5b^2 = 2$ where $\gamma = a + b\sqrt{-5}$ with $a, b \in \mathbb{Z}$. But this cannot happen. So $N(\gamma) = 1$, and hence $\gamma\overline{\gamma} = 1$. So $\gamma \in \mathcal{O}^\times_{\mathbb{Q}(\sqrt{-5})}$, and hence $(\gamma) = \mathcal{O}_{\mathbb{Q}(\sqrt{-5})}$. In particular $1 = 2(c + d\sqrt{-5}) + (1 + \sqrt{-5})(e + f\sqrt{-5})$ for some $c, d, e, f \in \mathbb{Z}$. Hence

$$1 = (2c + e - 5f) + (2d + e + f)\sqrt{-5}.$$

So $2c + e - 5f = 1$ and $2d + e + f = 0$. Taking the sum we get $2(c + d + e - 2f) = 1$. This cannot happen.

[13] The terminology is somewhat unfortunate.

23. Extra Topic: Transcendence of e and π

In this section, we take a glance at the transcendence theory. A first attempt is to prove that a certain number is not algebraic. More advanced topics will be to understand the relations between different transcendental numbers. Transcendence theory has many applications, and (perhaps) surprisingly it has been proved to be very useful to study algebraic objects.[14]

One useful technic for transcendence theory is the construction of appropriate auxiliary functions. The construction varies case by case, but there are some general tools, for example the Pigeonhole Principle, Geometry of Numbers and Siegel's Lemma.

In this section, we will present a proof of the transcendence of e and π.

The auxiliary function we shall use is the following one:

$$I(t) = \int_0^t e^{t-u} f(u) du,$$

where $f(x) \in \mathbb{C}[x]$. Denote by $n = \deg(f)$, then

$$I(t) = e^t \sum_{j=0}^{n} f^{(j)}(0) - \sum_{j=0}^{n} f^{(j)}(t) \tag{23.1}$$

where $f^{(j)}$ means the j-th derivative of the polynomial f. Here the computation is done by integration by parts:

$$\begin{aligned} I(t) &= \int_0^t (\frac{d}{du}(-e^{t-u}))f(u)du \\ &= (-e^{t-u}f(u))|_0^t - \int_0^t -e^{t-u}f'(u)du \\ &= e^t f(0) - f(t) + \int_0^t e^{t-u}f'(u)du \\ &= e^t f(0) - f(t) + e^t f'(0) - f'(t) + \int_0^t e^{t-u}f''(u)du \\ &= \cdots \\ &= e^t \sum_{j=0}^{n} f^{(j)}(0) - \sum_{j=0}^{n} f^{(j)}(t) \end{aligned}$$

We shall use the following notation. For any $f(x) = \sum_{j=0}^n a_j x^j$, set

$$f^\dagger(x) = \sum_{j=0}^{n} |a_j| x^j.$$

Then we have

$$|I(t)| \le |\int_0^t |e^{t-u}f(u)|du| \le |t| \max_u\{|e^{t-u}|\} \max_u\{|f(u)|\} \le |t|e^{|t|}|f^\dagger(|t|)| \tag{23.2}$$

Theorem 23.1. *The number e is transcendental.*

[14] In this application, transcendence theory is often combined with the height theory.

Proof. Assume e is an algebraic number, then it is a root of some

$$g(x) = b_r x^r + \ldots + b_0 \in \mathbb{Z}[x].$$

We may assume $b_0 \neq 0$. Let p be a prime number such that $p > \max(r, |b_0|)$. Define

$$f(x) = x^{p-1}(x-1)^p(x-2)^p \cdots (x-r)^p \in \mathbb{Z}[x].$$

Consider

$$J = b_0 I(0) + b_1 I(1) + \ldots + b_r I(r) \in \mathbb{Z}.$$

Let $n = \deg(f) = (r+1)p - 1$. By (23.1), we have

$$\begin{aligned} J &= b_0(1 - \sum_{j=0}^{n} f^{(j)}(0)) + b_1(e - \sum_{j=0}^{n} f^{(j)}(1)) + \ldots + b_r(e^r - \sum_{j=0}^{n} f^{(j)}(r)) \\ &= g(e) - b_0 \sum_{j=0}^{n} f^{(j)}(0) - \sum_{k=1}^{r} \sum_{j=0}^{n} b_k f^{(j)}(k) \\ &= -b_0 \sum_{j=p-1}^{n} f^{(j)}(0) - \sum_{k=1}^{r} \sum_{j=p}^{n} b_k f^{(j)}(k). \end{aligned}$$

Each term in this sum is divisible by $p!$ except for

$$f^{p-1}(0) = (p-1)!(-1)^{rp}(r!)^p.$$

Here we have used the fact that $p > r$. Hence $(p-1)!|J$, and $p \nmid J$ (since $p > |b_0|$). Now $p \nmid J$ implies that $J \neq 0$. Hence $(p-1)!|J$ implies that

$$|J| \geq (p-1)!. \tag{23.3}$$

On the other hand,

$$f^{\dagger}(j) = j^{p-1}(j+1)^p(j+2)^p \cdots (j+r)^p \leq (2r)^n$$

for any $0 \leq j \leq r$, we have by (23.2)

$$|J| \leq \sum_{j=0}^{r} |b_j||I(j)| \leq \sum_{j=0}^{r} |b_j| j e^j f^{\dagger}(j) \leq c(2r)^n \leq c\left((2r)^{r+1}\right)^p, \tag{23.4}$$

where $c = \max_{0 \leq j \leq r}(|b_j|)re^r$ independent of p.

Combining (23.3) and (23.4), we get

$$(p-1)! \leq c\left((2r)^{r+1}\right)^p$$

for every prime number $p > \max(r, |b_0|)$. In this inequality c and r are independent of p. But then this inequality cannot hold for p large enough. Hence we get a contradiction. □

Next we turn to π. Before moving on, let us prove the following lemma.

Lemma 23.2. *Let α be an algebraic number with minimal polynomial $f \in \mathbb{Q}[x]$. Let $c \in \mathbb{Z}$ be such that $cf \in \mathbb{Z}[x]$. Then $c\alpha$ is an algebraic integer.*

Proof. Write $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0 \in \mathbb{Q}[x]$. Then $ca_i \in \mathbb{Z}$ for all $i \in \{0, \ldots, n-1\}$. Hence

$$g(x) = x^n + ca_{n-1}x^{n-1} + c^2a_{n-2}x^{n-2} + \ldots + c^n a_0 \in \mathbb{Z}[x].$$

But $g(c\alpha) = c^n f(\alpha) = 0$. Hence $c\alpha$ is an algebraic integer. □

Theorem 23.3. *The number π is transcendental.*

Proof. Suppose π is algebraic. Then $\theta = i\pi$ is algebraic by Proposition 22.9. Let $f \in \mathbb{Q}[x]$ be the minimal polynomial of θ, and let $\theta_1 = \theta$, $\theta_2, \ldots, \theta_r$ be the roots of f in $\mathbb{C}$. Let $b \in \mathbb{Z}$ be positive such that $bf \in \mathbb{Z}[x]$. Then $b\theta_j$ is an algebraic integer for any $j \in \{1, \ldots, r\}$ by Lemma 23.2. Since $e^{i\pi} = -1$, we have

$$(1 + e^{\theta_1}) \cdots (1 + e^{\theta_r}) = 0.$$

Expanding the left hand side, we get 2^r terms of the form e^{ϕ} where $\phi = \epsilon_1\theta_1 + \ldots + \epsilon_r\theta_r$ with $\epsilon_j \in \{0, 1\}$ for all $j \in \{1, \ldots, r\}$. The number ϕ may or may not be zero.

Let $\phi_1, \ldots, \phi_m$ be the non-zero expressions of this form, then the remaining $2^r - m$ ϕ's are 0. Hence we get

$$(2^r - m) + e^{\phi_1} + \ldots + e^{\phi_m} = 0.$$

Let us look at the polynomial $\prod_\phi (x - \phi)$, which has degree 2^r. It is symmetric in $\theta_1, \ldots, \theta_r$, meaning that this polynomial does not change if we reorder the $\theta_1, \ldots, \theta_r$. Hence this polynomial is in $\mathbb{Q}[x]$.[15] But $\prod_\phi (x - \phi) = x^{2^r - m} \prod_{j=1}^m (x - \phi_i)$, and hence $\prod_{j=1}^m (x - \phi_i) \in \mathbb{Q}[x]$.

Let p be a prime number such that $p > b^m|\phi_1 \cdots \phi_m|$. Define

$$f(x) = b^{mp}x^{p-1}(x - \phi_1)^p \cdots (x - \phi_m)^p.$$

Then $\deg(f) = n = (m+1)p - 1$ and $f \in \mathbb{Q}[x]$. Moreover $b\phi_j$ is an algebraic integer for each $j \in \{1, \ldots, m\}$ since $b\theta_j$ is an algebraic integer for each $j \in \{1, \ldots, r\}$. So every coefficient of f is an algebraic integer by Proposition 22.9. Hence $f \in \mathbb{Z}[x]$ by Lemma 22.6. Define

$$J = I(\phi_1) + \ldots I(\phi_m).$$

By (23.1), we have

$$J = -(2^r - m) \sum_{j=0}^{n} f^{(j)}(0) - \sum_{j=0}^{n} \sum_{k=1}^{m} f^{(j)}(\phi_k).$$

Observe that the sum over k is a symmetric polynomial in $b\phi_1, \ldots, b\phi_m$ with integer coefficients and hence a symmetric polynomial with integer coefficients in the 2^r numbers $b\phi = b(\epsilon_1\theta_1 + \ldots + \epsilon_r\theta_r)$. Hence again this sum is an algebraic integer in $\mathbb{Q}$, and hence in $\mathbb{Z}$ by Lemma 22.6.

Since $f^{(j)}(\phi_k) = 0$ for all $j < p$, we deduce that the double sum in the expression for J is an integer divisible by $p!$. Let us study the other sum in the expression for J. First observe that $f^{(j)}(0) = 0$ for all $j < p - 1$ and $f^{(j)}(0)$ is divisible by $p!$ for all $j \geq p$. Now we want to understand $f^{(p-1)}(0)$. We have

$$f^{(p-1)}(0) = b^{mp}(-1)^{mp}(p-1)!(\phi_1 \cdots \phi_m)^p.$$

Recall that $f \in \mathbb{Z}[x]$. Looking at the constant term of f, we get that $b^{mp}(-1)^{mp}(\phi_1 \cdots \phi_m)^p \in \mathbb{Z}$. Hence $f^{(p-1)}(0) \in \mathbb{Z}$ is divisible by $(p-1)!$. Since $p > b^m|\phi_1 \cdots \phi_m|$, we have $p \nmid f^{(p-1)}(0)$.

[15] This can be seen as a consequence of the fundamental theorem of elementary symmetric functions. It can also be easily proven if one has some basic knowledge of the Galois group.

To sum it up, $(p-1)!|J$ and $p \nmid J$. Thus $J \neq 0$, and

$$|J| \geq (p-1)!.$$

On the other hand (23.2) implies

$$|J| \leq \sum_{k=1}^{m} |\phi_k| e^{|\phi_k|} f^{\dagger}(|\phi_k|) = \sum_{k=1}^{m} |\phi_k| e^{|\phi_k|} b^{mp} |\phi_k|^{p-1} (|\phi_k| + |\phi_1|)^p \cdots (|\phi_k| + |\phi_m|)^p \leq c_1 c_2^p$$

for some numbers c_1, c_2 independent of p.[16] Now we get

$$(p-1)! \leq c_1 c_2^p$$

for all $p > b^m |\phi_1 \cdots \phi_m|$. But this cannot be true for p large enough. Hence we get a contradiction. □

[16]Of course it is possible to write down the expressions for c_1, c_2 in terms of b, m and the ϕ_j's. But since all we want is that these two numbers are independent of p, we may as well simply introduce the new symbols c_1, c_2 and say that they are independent of p.