Algebraic Number Theory

Joseph R. Mileti

May 11, 2012

Contents

Chapter 1

Introduction

1.1 Sums of Squares

Question 1.1.1. *Which numbers can be written as the sum of two squares? As we will eventually see, solving this problem boils down to determining which primes can be written as the sum of two squares. For example,* $5 = 1 + 4$*,* $13 = 4 + 9$*,* $61 = 25 + 36$*, etc.*

We first start with a simple necessary condition.

Proposition 1.1.2. *If* p *is an odd prime which can be written as the sum of two squares, then* $p \equiv 1 \pmod 4$*.*

Proof. It is straightforward to check that for all $n \in \mathbb{Z}$, either $n^2 \equiv 0 \pmod 4$ or $n^2 \equiv 1 \pmod 4$. Thus, the sum of two squares must equal one of 0, 1, or 2 modulo 4, and so can not be 3 modulo 4. Now every odd number is congruent to either 1 or 3 modulo 4, so if p is an odd prime which can be written as the sum of two squares, then $p \equiv 1 \pmod 4$. □

Suppose that p is a prime which is the sum of two squares. We can then fix $a, b \in \mathbb{Z}$ with $p = a^2 + b^2$. In the Gaussian integers $\mathbb{Z}[i]$ we have

$$p = a^2 + b^2 = (a + bi)(a - bi)$$

so the number p, which is prime in $\mathbb{Z}$, factors in an interesting manner over the larger ring $\mathbb{Z}[i]$. There is a converse to this as well. In fact, we will eventually be able to show the following theorem.

Theorem 1.1.3. *Let* $p \in \mathbb{Z}$ *be an odd prime. The following are equivalent.*

1. *There exist* $a, b \in \mathbb{Z}$ *with* $p = a^2 + b^2$*.*
2. p *is reducible (and hence no longer prime) in* $\mathbb{Z}[i]$*.*
3. -1 *is a square modulo* p*, i.e. there exists* $x \in \mathbb{Z}$ *such that* $x^2 \equiv -1 \pmod p$*.*
4. $p \equiv 1 \pmod 4$*.*

Putting the above information together, it follows that p can be written as the sum of two squares exactly when p (which is prime/irreducible in $\mathbb{Z}$) fails to be irreducible in the larger ring $\mathbb{Z}[i]$. We have turned a number-theoretic question into one about factorizations in a new ring. In order to take this perspective, we need a solid understanding of factorizations in the ring $\mathbb{Z}[i]$. For example, is the ring a UFD so that we have unique factorizations into irreducibles? If not, how badly does factorization break down?

Moreover, the above theorem establishes a connection with the squares in the ring $\mathbb{Z}/p\mathbb{Z}$. Determining which elements in these rings are squares is a fascinating problem and leads to the beautiful result known as Quadratic Reciprocity. Thus, from this simple example, we see how basic number-theoretic questions can be understood and hopefully solved using the perspective of the algebraic objects you studied in Abstract Algebra. This course is largely devoted to justifying this claim.

1.2 Pythagorean Triples

Suppose that you want to understand all Pythagorean triples, that is triples (a, b, c) of positive integers with $a^2 + b^2 = c^2$. To find these, it suffices to find all so-called primitive Pythagorean triples, that is Pythagorean triples (a, b, c) with $\gcd(a, b, c) = 1$, because a general Pythagorean triple is an integer multiple of a primitive ones.. These are called primitive triples. Given a primitive triple (a, b, c), we have

$$c^2 = a^2 + b^2 = (a + bi)(a - bi)$$

Now one can show (and we will do this), that gcd's make sense in $\mathbb{Z}[i]$ and that $a+bi$ and $a-bi$ are relatively prime assuming that (a, b, c) is primitive. We will then be able to show that if the product of two relatively prime elements of $\mathbb{Z}[i]$ is a square, then each of the factors must be a square. In particular, we can fix $m, n \in \mathbb{Z}$ with

$$a + bi = (n + mi)^2$$

We then have

$$a + bi = (n^2 - m^2) + 2mn \cdot i$$

We conclude that $a = n^2 - m^2$ and $b = 2mn$. From here, it is easy to show that $c = m^2 + n^2$. The converse also holds (any triple (a, b, c) generated this way is a Pythagorean triple), so we obtain a way to parametrize all primitive Pythagorean triples. There are other more elementary ways to derive these parameterizations (and we will see them), but this method is faster, generalizes better, and "explains" the formulas in a more satisfying fashion.

1.3 Solving Other Diophantine Equations

Suppose that we try to find all integer solutions to

$$x^3 = y^2 + 1$$

The solution $(1, 0)$ is clear, but are there any others? Notice that if y is odd, then the right-hand side is congruent to 2 modulo 4, but 2 is not a cube modulo 4. Thus, we can assume that y is even. We can factor the right-hand side as

$$x^3 = (y + i)(y - i)$$

We will be able to show that $y + i$ and $y - i$ are relatively prime in $\mathbb{Z}[i]$, so both are cubes. In particular, we can write

$$y + i = (a + bi)^3 = (a^3 - 3ab^2) + (3a^2b - b^3)i$$

so

$$y = a(a^2 - 3b^2) \qquad 1 = b(3a^2 - b^2)$$

The right-hand equation implies $b = \pm 1$. If $b = 1$, we get $1 = 3a^2 - 1$, so $2 = 3a^2$, a contradiction. If $b = -1$, we get $1 = -(3a^2 - 1)$, so $-3a^2 = 0$ and $a = 0$. Thus, we conclude $a = 0$ and $b = -1$. Therefore, $x = 1$ and this gives $y = 0$.

Suppose that we try to find all integer solutions to

$$x^3 = y^2 + 19$$

If you follow the above, you might try to factor this as

$$x^3 = (y + \sqrt{-19})(y - \sqrt{-19})$$

where we are working in the ring $\mathbb{Z}[\sqrt{-19}]$. With some work, one can show that if (x, y) is a solution, then the only common divisors of $y + \sqrt{-19}$ and $y - \sqrt{-19}$ are ± 1, so they are relatively prime. As above, one then hopes that each of the factors on the right are cubes, and working out similar but more complicated computations would lead one to conclude that there are no solutions. This would all be great except for the fact that

$$18^2 + 19 = 343 = 7^3$$

so $(18, 7)$ is a solution. There is something different about the ring $\mathbb{Z}[\sqrt{-19}] = \{a + b\sqrt{-19} : a, b \in \mathbb{Z}\}$ which makes this argument fail, and the fundamental fact is that $\mathbb{Z}[\sqrt{-19}]$ is not a UFD. However, it turns out that a slightly larger ring is a UFD, and using this one can show that the only solutions are $(\pm 18, 7)$. This illustrates how the ring-theoretic structure of certain generalizations of $\mathbb{Z}$ have implications for $\mathbb{Z}$ itself.

1.4 Fermat's Last Theorem

We know that there exist nontrivial solutions to $x^2 + y^2 = z^2$. Fermat's Last Theorem is the statement that if $n \geq 3$, then there are no solutions to

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

with each of x, y, z positive integers. Fermat scribbled a note in the margin of one of his books stating that he had a proof but the margin was too small to contain it. For centuries, mathematicians attempted to prove this result. If there exists a nontrivial solution for some n, then some straightforward calculations show that there must be a nontrivial solution for either $n = 4$ or for some odd prime p. Fermat did show that

$$x^4 + y^4 = z^4$$

has no nontrivial solutions. Suppose then that p is an odd prime and we want to show that

$$z^p = x^p + y^p$$

has no nontrivial solution. The idea is to factor the right-hand side. Although it may not be obvious at this point, by setting $\zeta = e^{2\pi i/p}$, it turns out that

$$x^p + y^p = (x + y)(x + \zeta y)(x + \zeta^2 y) \cdots (x + \zeta^{p-1} y)$$

Thus, if (x, y, z) is a nontrivial solution, we have

$$z^p = (x + y)(x + \zeta y)(x + \zeta^2 y) \cdots (x + \zeta^{p-1} y)$$

The idea then is to work in the ring

$$\mathbb{Z}[\zeta] = \{a_0 + a_1\zeta + a_2\zeta^2 + \cdots + a_{p-1}\zeta^{p-1} : a_i \in \mathbb{Z}\}$$

Again, one can show that if (x, y, z) is a nontrivial solution, then the factors on the right are "relatively prime" in $\mathbb{Z}[\zeta]$. Lamé put forward this argument and claimed it implied that the factors on the right must then be p^{th} powers, from which he derived a contradiction and hence claimed a proof of Fermat's Last

Theorem. Liouville pointed out that this argument relied essentially on the ring $\mathbb{Z}[\zeta]$ being a UFD (though he did not have that terminology). There do exist rings where $\mathbb{Z}[\zeta]$ does not have unique factorization, but this was major progress.

In an attempt to "fix" this lack of unique factorization and also to pursue generalizations of Quadratic Reciprocity, Kummer introduced so-called "ideal numbers" which can be used to restore unique factorization in these types of rings. These "ideal numbers" were later abstracted and generalized by Dedekind into the ideals of ring theory. Thus, investigations in number theory itself led to some fundamental concepts in abstract algebra. It turns out that by moving from elements to ideals, one restores a certain type of unique factorization.

1.5 The Fundamental Theorem of Arithmetic

Throughout this section, the key ring-theoretic fact about the integers that we will use repeatedly is that an integer is prime if and only if it is irreducible. Recall that primes are irreducible in every integral domain, but the converse is not true (and we will certainly see examples of this throughout the course). However, in this section, pay careful attention to when we are using using the stronger property of being prime (rather than just irreducible).

Proposition 1.5.1. *Every nonzero nonunit $n \in \mathbb{Z}$ is a product of primes.*

Proof. We first prove the result for $n \in \mathbb{N}$ by strong induction. If $n = 2$, we are done because 2 itself is prime. Suppose that $n > 2$ and we have proven the result for all k with $1 < k < n$. If n is prime, we are done. Suppose that n is not prime and fix a divisor $c \mid n$ with $1 < c < n$. Fix $d \in \mathbb{N}$ with $cd = n$. We then have that $1 < d < n$, so by induction, both c and d are products of primes, say $c = p_1 p_2 \cdots p_k$ and $d = q_1 q_2 \cdots q_\ell$ with each p_i and q_j prime. We then have

$$n = cd = p_1 p_2 \cdots p_k q_1 q_2 \cdots q_\ell$$

so n is a product of primes. The result follows for $n \in \mathbb{N}$ follows by induction.

Suppose now that $n \in \mathbb{Z}$ is negative. We then have that $-n \in\in \mathbb{N}$ and $-n \geq 2$. Therefore, from above, we may write $n = p_1 p_2 \cdots p_k$ where the p_i are prime. We then have

$$-n = (-p_1) p_2 \cdots p_k$$

Since $-p_1$ is also prime, the result follows. □

Definition 1.5.2. *Let $p \in \mathbb{N}^+$ be prime. Define a function $ord_p \colon \mathbb{Z} \to \mathbb{N} \cup \{\infty\}$ as follows. Let $ord_p(0) = \infty$, and given $a \in \mathbb{Z} - \{0\}$, let $ord_p(a)$ be the largest $k \in \mathbb{N}$ such that $p^k \mid a$.*

Lemma 1.5.3. *Let $p \in \mathbb{N}^+$ be prime, let $a \in \mathbb{Z}$, and let $k \in \mathbb{N}$. The following are equivalent.*

1. *$ord_p(a) = k$*
2. *$p^k \mid a$ and $p^{k+1} \nmid a$*
3. *There exists $m \in \mathbb{Z}$ with $a = p^k m$ and $p \nmid m$*

Proof.

- $1 \to 2$ is immediate.
- $2 \to 1$: Suppose that $p^k \mid a$ and $p^{k+1} \nmid a$. We clearly have $ord_p(a) \geq k$. Suppose that there exists $\ell > k$ with $p^\ell \mid a$. Since $\ell > k$, we have $\ell \geq k + 1$. This implies that $p^{k+1} \mid p^\ell$, so since $p^\ell \mid a$ we conclude that $p^{k+1} \mid a$. This contradicts our assumption. Therefore, there is no $\ell > k$ with $p^\ell \mid a$, and hence $ord_p(a) = k$.

- $2 \to 3$: Suppose that $p^k \mid a$ and $p^{k+1} \nmid a$. Fix $m \in \mathbb{Z}$ with $a = p^k m$. If $p \mid m$, then we may fix $n \in \mathbb{Z}$ with $m = pn$, which would imply that $a = p^k pn = p^{k+1}n$ contradicting the fact that $p^{k+1} \nmid a$. Therefore, we must have $p \nmid m$.
- $3 \to 2$: Fix $m \in \mathbb{Z}$ with $a = p^k m$ and $p \nmid m$. We clearly have $p^k \mid a$. Suppose that $p^{k+1} \mid a$ and fix $n \in \mathbb{Z}$ with $a = p^{k+1}n$. We then have $p^k m = p^{k+1}n$, so $m = pn$. This implies that $p \mid m$, which is a contradiction. Therefore, $p^{k+1} \nmid a$.

□

Theorem 1.5.4. *Let $p \in \mathbb{N}^+$ be prime. We have the following.*

1. *$ord_p(ab) = ord_p(a) + ord_p(b)$ for all $a, b \in \mathbb{Z}$.*
2. *$ord_p(a^n) = n \cdot ord_p(a)$ for all $a \in \mathbb{Z}$ and $n \in \mathbb{N}^+$.*
3. *$ord_p(a+b) \geq \min\{ord_p(a), ord_p(b)\}$ for all $a, b \in \mathbb{Z}$.*
4. *$ord_p(a+b) = \min\{ord_p(a), ord_p(b)\}$ for all $a, b \in \mathbb{Z}$ with $ord_p(a) \neq ord_p(b)$.*

Proof. See Homework 1. □

Lemma 1.5.5. *Let $p \in \mathbb{Z}$ be prime.*

1. *For any prime q that is an associate of p, we have $ord_p(q) = 1$.*
2. *For any prime q that is not an associate of p, we have $ord_p(q) = 0$.*
3. *For any unit u, we have $ord_p(u) = 0$.*

Proof.
1. Suppose that q is a prime that is an associate of p. Fix a unit u with $q = pu$. Notice that if $p \mid u$, then since $u \mid 1$, we conclude that $p \mid 1$, which would imply that p is a unit. Since p is not a unit, it follows that $p \nmid u$. Therefore, $ord_p(q) = 1$ by Lemma 1.5.3.
2. Suppose that q is a prime that is not an associate of p. Since q is prime, it is irreducible, so its only divisors are units and associates. Since p is not a unit nor an associate of q, it follows that $p \nmid q$. Therefore, $ord_p(q) = 0$.
3. This is immediate because if $p \mid u$, then since $u \mid 1$, we could conclude that $p \mid 1$. This implies that p is a unit, which is a contradiction.

□

Lemma 1.5.6. *Let $n \in \mathbb{Z}$ with $n \neq 0$. Letting $\mathbb{P}$ be the set of primes, we have that*

$$\{p \in \mathbb{P} : ord_p(n) > 0\}$$

is finite.

Proof. If $ord_p(n) > 0$, then $p \mid n$, hence $p \leq |n|$ because $n \neq 0$. The result follows from the fact that for any nonzero n, the set $\{m \in \mathbb{Z} : m \leq |n|\}$ is finite. □

Lemma 1.5.7. *Let $n \in \mathbb{Z}$ and let $p \in \mathbb{N}^+$ be prime. Suppose that u is a unit, that q_i are primes, and that*

$$n = uq_1q_2 \cdots q_k$$

We then have that exactly $ord_p(n)$ many of the q_i are associates of p.

Proof. Since

$$n = uq_1q_2\cdots q_k$$

we have

$$\begin{aligned} ord_p(n) &= ord_p(uq_1q_2\cdots q_k) \\ &= ord_p(u) + \sum_{i=1}^{k} ord_p(q_i) \\ &= \sum_{i=1}^{k} ord_p(q_i) \end{aligned}$$

The terms on the right are 1 when q_i is an associate of p and 0 otherwise. The result follows. □

Theorem 1.5.8 (Fundamental Theorem of Arithmetic)**.** *Every nonzero nonunit $n \in \mathbb{Z}$ factors uniquely into a product of primes up to order and associates. In other words, suppose that $n \notin \{-1, 0, 1\}$ and that*

$$uq_1q_2\cdots q_\ell = n = wr_1r_2\cdots r_\ell$$

where u and w are units, and each of the q_i and r_j are primes. We then have that $k = \ell$ and there exists $\sigma \in S_k$ such that q_i and $r_{\sigma(i)}$ are associates for all i.

Proof. Let p be an arbitrary prime. We know from the lemma that exactly $ord_p(n)$ many of the q_i are associates of p, and also that exactly $ord_p(n)$ many of the r_j are associates of p. Thus, for every prime p, there are an equal number of associates of p on each side. Matching up the elements on the left with corresponding associates on the right gives the required permutation. □

Proposition 1.5.9. *Let $m, n \in \mathbb{Z}$. The following are equivalent.*

1. *m and n are associates.*
2. *$ord_p(m) = ord_p(n)$ for all primes p.*

Proof. Suppose first that m and n are associates. Fix a unit u with $m = nu$. For any prime p, we then have

$$\begin{aligned} ord_p(m) &= ord_p(nu) \\ &= ord_p(n) + ord_p(u) \\ &= ord_p(n) \end{aligned}$$

Suppose conversely that $ord_p(m) = ord_p(n)$ for all primes p. Notice that $k = 0$ if and only if $ord_p(k) = \infty$ for all primes p. Also, notice that k is a unit if and only if $ord_p(k) = 0$ for all primes p because every nonzero nonunit is a product of primes. Thus, the result holds if either of m or n (or both) are 0 or units. Assume then that both m and n are nonzero nonunits. Write m and n as products of primes, say

$$m = q_1q_2\cdots q_k$$

and

$$n = r_1r_2\cdots r_\ell$$

Now using the above lemma, for any prime p, we have that exactly $ord_p(m)$ many of the q_i are associates of p, and exactly $ord_p(n)$ many of the r_j are associates of p. Since $ord_p(m) = ord_p(n)$ for all primes p, we conclude that for any prime p, the number of associates of p amongst the q_i equals the number of associates of p amongst the r_j. Thus, $k = \ell$ and we may match up associate pairs in the two factorizations. By

rearranging the products, we may assume that q_i are r_i are associates for every i. For each i, fix a unit u with $r_i = u_i q_i$. We then have

$$\begin{aligned} n &= r_1 r_2 \cdots r_k \\ &= (u_1 q_1)(u_2 q_2) \cdots (u_k q_k) \\ &= u_1 u_2 \cdots u_k \cdot q_1 q_2 \cdots q_k \\ &= (u_1 u_2 \cdots u_k) m \end{aligned}$$

Since the product of units is a unit, we conclude that $u_1 u_2 \cdots u_k$ is a unit. Therefore, m and n are associates. □

Proposition 1.5.10. *Let $d, n \in \mathbb{Z}$. We have that $d \mid n$ if and only if $ord_p(d) \leq ord_p(n)$ for all primes p.*

Proof. Suppose that $d \mid n$. Fix $k \in \mathbb{Z}$ with $n = dk$. For any prime p, we have $ord_p(n) = ord_p(d) + ord_p(k)$. Since $ord_p(k) \geq 0$, we conclude that $ord_p(d) \leq ord_p(n)$ for any prime p.

Suppose conversely that $ord_p(d) \leq ord_p(n)$ for all primes p. Let $F = \{p \in \mathbb{P} : ord_p(n) > 0\}$. We know from above that F is finite, so we may let

$$k = \prod_{p \in F} p^{ord_p(n) - ord_p(d)}$$

For any $p \in F$, we have $ord_p(k) = ord_p(n) - ord_p(d)$ and therefore

$$\begin{aligned} ord_p(dk) &= ord_p(d) + ord_p(k) \\ &= ord_p(d) + (ord_p(n) - ord_p(d)) \\ &= ord_p(n) \end{aligned}$$

Also, for any prime $p \in \mathbb{P} \backslash F$, we have $ord_p(d) \leq ord_p(n) = 0$, so

$$\begin{aligned} ord_p(dk) &= ord_p(d) + ord_p(k) \\ &= 0 + 0 \\ &= ord_p(n) \end{aligned}$$

It follows that n and dk are associates. In particular, we have that $dk \mid n$. Since $d \mid dk$, we conclude that $d \mid n$. □

Proposition 1.5.11. *Let $m, n \in \mathbb{Z}$. We have $\gcd(m, n) = 1$ if and only if for all primes p, at most one of $ord_p(m)$ or $ord_p(n)$ is nonzero.*

Proof. Suppose there exists a prime p such that both $ord_p(m) > 0$ and $ord_p(n) > 0$. Fix such a prime p. We then have that p a common divisor of m and n, so $\gcd(m, n) \neq 1$.

Conversely, suppose that $\gcd(m, n) \neq 1$. Let $d = \gcd(m, n)$. Since $d > 1$, we may fix a prime $p \mid d$. We then have that p is a common divisor of m and n, so $ord_p(m) \geq 1$ and $ord_p(n) \geq 1$. □

Proposition 1.5.12. *Let $m \in \mathbb{Z}$ and let $n \in \mathbb{N}^+$. We have that some associate of m is an n^{th} power in $\mathbb{Z}$ if and only if $n \mid ord_p(m)$ for all primes p.*

Proof. Suppose first that some associate of m is an n^{th} power in $\mathbb{Z}$. We may then fix $u, d \in \mathbb{Z}$ such that u is a unit and $mu = d^n$. For any prime p, we then have that

$$\begin{aligned} ord_p(m) &= ord_p(u^{-1} d^n) \\ &= ord_p(u^{-1}) + n \cdot ord_p(d) \\ &= n \cdot ord_p(d) \end{aligned}$$

Therefore, $n \mid ord_p(m)$ for every prime p.

Suppose conversely that $n \mid ord_p(m)$ for all primes p. Let $F = \{p \in \mathbb{P} : ord_p(m) > 0\}$ and recall that F is finite. For each $p \in F$, fix $e_p \in \mathbb{Z}$ with $ord_p(m) = ne_p$. Notice that $e_p > 0$ for all $p \in F$ because $n > 0$ and each $ord_p(m) > 0$. Define

$$d = \prod_{p \in F} p^{e_p}$$

We then have

$$d^n = \prod_{p \in F} p^{ne_p}$$

hence $ord_p(d^n) = ne_p = ord_p(m)$ for all primes p. Therefore, m and d^n are associates. □

Corollary 1.5.13. *Let $m \in \mathbb{Z}$ be nonzero and let $n \in \mathbb{N}^+$.*

1. *Suppose that n is odd. We then have that m is an n^{th} power in $\mathbb{Z}$ if and only if $n \mid ord_p(m)$ for all primes p.*
2. *Suppose that n is even. We then have that m is an n^{th} power in $\mathbb{Z}$ if and only if $m \geq 0$ and $n \mid ord_p(m)$ for all primes p.*

Proof. The first statement follows from the fact that if n is odd, then each unit of $\mathbb{Z}$ (i.e. each of ± 1) is a n^{th} power because $1^n = 1$ and $(-1)^n = -1$. Therefore, if some associate of m is an n^{th} power, then m itself must also be an n^{th} power.

The second statement follows from the fact that ± 1 are the only units and the fact that if n is even, then all n^{th} powers are positive. Thus, if m is an n^{th} power, then $m \geq 0$. Conversely, if $m \geq 0$ and some associate of m is an n^{th} power, then that associate must be m itself because the only associates of m are $\pm m$. □

Theorem 1.5.14. *Suppose that $a, b \in \mathbb{Z}$ are relatively prime and that ab is a square. If both $a \geq 0$ and $b \geq 0$, then both a and b are squares.*

Proof 1. Suppose that both $a \geq 0$ and $b \geq 0$. Since ab is a square, we know that $2 \mid ord_p(ab)$ for all primes p, hence $2 \mid ord_p(a) + ord_p(b)$ for all primes p. Furthermore, since a and b are relatively prime, we know that for each prime p, at most one of $ord_p(a)$ or $ord_p(b)$ is nonzero. Let p be prime. If $ord_p(a) = 0$, then trivially $2 \mid ord_p(a)$. If $ord_p(a) \neq 0$, then $ord_p(b) = 0$, hence $ord_p(a) = ord_p(a) + ord_p(b)$ and so $2 \mid ord_p(a)$. Therefore, $2 \mid ord_p(a)$ for all p. Since $a \geq 0$, we conclude that a is a square. The proof that b is a square is completely analogous (or simply note that $ab = ba$). □

Proof 2. First suppose that $a = 0$. Since a and b are relatively prime, this implies that $b = 1$, so clearly both a and b are squares. Similarly, if $b = 0$, then $a = 1$ and we are done.

Now suppose that $a = 1$. We then have that $b = ab$ so since ab is a square we clearly have that b is a square. Similarly, if $b = 1$, then $a = ab$ is a square.

Suppose then that $a, b \geq 2$. Write each of a and b in terms of its unique prime factorization:

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$$
$$b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}$$

where we are assume that the p_i are distinct primes and $\alpha_i, \beta_i, \gamma_i \geq 0$ for all i (although some may be zero). We then have

$$ab = p_1^{\alpha_1 + \beta_1} p_2^{\alpha_2 + \beta_2} \cdots p_k^{\alpha_k + \beta_k}$$

Now ab is a square, so by Proposition 1.44 in the notes we conclude that $\alpha_i + \beta_i$ is even for all i.

We now show that a is a square. Fix i with $1 \leq i \leq k$. If $\alpha_i = 0$, then α_i is certainly even. Suppose that $\alpha_i > 0$. We must have $\beta_i = 0$ because otherwise p_i would divide both a and b, which would contradict the fact that a and b are relatively prime. Therefore, $\alpha_i = \alpha_i + \beta_i$ is even. We have shown that α_i is even for all i, so a is a square. The proof that b is a square is completely analogous (or simply note that $ab = ba$). □

Notice that we needed to assume that $a, b \geq 0$. To see why this is true, simply notice that $6 = (-2)\cdot(-3)$, but -2 and -3 are not squares in $\mathbb{Z}$.

1.6 Primes

Proposition 1.6.1. *There are infinitely many primes.*

Proof. We know that 2 is a prime, so there is at least one prime. We will take an arbitrary given finite list of primes and show that there exists a prime which is omitted. Suppose then that $p_1, p_2, \ldots, p_k$ is an arbitrary finite list of prime numbers with $k \geq 1$. We show that there exists a prime not in the list. Let

$$n = p_1 p_2 \cdots p_k + 1$$

We have $n \geq 3$, so by the above corollary we know that n is divisible by some prime q. If $q = p_i$, we would have that $q \mid n$ and also $q \mid p_1 p_2 \cdots p_k$, so $q \mid (n - p_1 p_2 \cdots p_k)$. This would imply that $q \mid 1$, a contradiction. Therefore $q \neq p_i$ for all i, and we have succeeded in finding a prime not in the list. □

Theorem 1.6.2. *There are infinitely many primes $p \equiv 3 \pmod 4$.*

Proof. We know that 3 is a prime, so there is at least one such prime. We will take an arbitrary given finite list of primes and show that there exists a prime which is omitted. Suppose then that $p_1, p_2, \ldots, p_k$ is an arbitrary finite list of prime numbers with $k \geq 1$ such that $p_i \equiv 3 \pmod 4$ for all i. We show that there exists a prime not in the list. Let

$$n = 4p_1 p_2 \cdots p_k - 1$$

Notice that $n \geq 4 - 1 > 3$, so n is a product of primes. Now $n \equiv -1 \equiv 3 \pmod 4$, so n is odd, and hence 2 can not appear in the product. Also, all primes other than 2 are odd, so all such primes must be congruent to one of 1 or 3 modulo 4.

Notice that if $a \equiv 1 \pmod 4$ and $b \equiv 1 \pmod 4$, then $ab \equiv 1 \pmod 4$. Since $n \equiv 3 \pmod 4$, it is impossible that all of these prime divisors of n are congruent to 1 modulo 3. We conclude that some prime q in the factorization of n satisfies $q \equiv 3 \pmod 4$. For this q, we have that $q \mid n$. Now suppose that $q = p_i$ for some i. We would then have that $q \mid n$ and $q \mid 4p_1 p_2 \cdots p_k$, hence $q \mid (4p_1 p_2 \cdots p_k - n)$, i.e. $q \mid 1$. This is contradiction, so it follows that $q \neq p_i$ for any i. We have thus found a prime q such that $q \equiv 3 \pmod 4$ and $q \neq p_i$ for all i, so q is a prime congruent to 3 modulo 4 which is not in the list of p_i. □

1.7 Pythagorean Triples from an Elementary Viewpoint

Definition 1.7.1. *A* Pythagorean triple *is a triple of positive integers (a, b, c) with $a^2 + b^2 = c^2$.*

Definition 1.7.2. *A Pythagorean triple (a, b, c) is* primitive *if $\gcd(a, b, c) = 1$.*

Proposition 1.7.3. *A Pythagorean triple (a, b, c) is primitive if and only if every pair of elements from (a, b, c) are relatively prime.*

Proof. Let (a, b, c) be a Pythagorean triple. If every pair of elements from (a, b, c) are relatively prime, then $\gcd(a, b) = 1$, so trivially $\gcd(a, b, c) = 1$ and hence (a, b, c) is primitive. Suppose conversely that some pair from (a, b, c) are not relatively prime. We have the following cases.

- Suppose that $\gcd(a, b) \neq 1$. Fix a prime p dividing both a and b. We then have that $p \mid a^2$ and $p \mid b^2$, so $p \mid (a^2 + b^2)$ which is to say that $p \mid c^2$. Since p is prime, we conclude that $p \mid c$. Therefore, p is a divisor of each of a, b, c, hence $\gcd(a, b, c) \neq 1$.

- Suppose that $\gcd(a,c) \neq 1$. Fix a prime p dividing both a and c. We then have that $p \mid a^2$ and $p \mid c^2$, so $p \mid (c^2 - a^2)$ which is to say that $p \mid b^2$. Since p is prime, we conclude that $p \mid b$. Therefore, p is a divisor of each of a, b, c, hence $\gcd(a,b,c) \neq 1$.
- Suppose that $\gcd(b,c) \neq 1$. Fix a prime p dividing both b and c. We then have that $p \mid b^2$ and $p \mid c^2$, so $p \mid (c^2 - b^2)$ which is to say that $p \mid a^2$. Since p is prime, we conclude that $p \mid a$. Therefore, p is a divisor of each of a, b, c, hence $\gcd(a,b,c) \neq 1$.

Thus, in all cases, we have that $\gcd(a,b,c) \neq 1$, so (a,b,c) is not primitive. □

Proposition 1.7.4. *Every Pythagorean triple is an integer multiple of a primitive Pythagorean triple.*

Proof. Let (a,b,c) be a Pythagorean triple. Let $d = \gcd(a,b,c)$. We then have

$$\left(\frac{a}{d}\right)^2 + \left(\frac{b}{d}\right)^2 = \frac{a^2+b^2}{d^2} = \frac{c^2}{d^2} = \left(\frac{c}{d}\right)^2$$

so $(a/d, b/d, c/d)$ is also a Pythagorean triple. If it is not primitive, then it is straightforward to argue that $d \neq \gcd(a,b,c)$. □

Proposition 1.7.5. *Let (a,b,c) be a primitive Pythagorean triple. We then have that exactly one of a or b is even, and also that c is odd.*

Proof. If both a and b are even, then $\gcd(a,b) \geq 2$, so from above we have that (a,b,c) is not primitive. Suppose that both a and b are odd. We then have that $a^2 \equiv 1 \pmod 4$ and $b^2 \equiv 1 \pmod 4$, so $a^2 + b^2 \equiv 2 \pmod 4$. This implies that $c^2 \equiv 2 \pmod 4$ which is a contradiction because the only squares modulo 4 are 0 and 1.

Since exactly one of a and b is even, we can't have that c is even for otherwise some pair of elements would not be relatively prime. □

Notice that if (a,b,c) is a Pythagorean triple, then trivially (b,a,c) is a Pythagorean triple. We now determine all primitive Pythagorean triples with b even.

Theorem 1.7.6. *Let (a,b,c) be a primitive Pythagorean triple with b even. There exist relatively prime positive integers $m < n$ having distinct parities (i.e. one even and one odd) such that*

$$a = n^2 - m^2 \qquad b = 2mn \qquad c = m^2 + n^2$$

Furthermore, every such triple is a primitive Pythagorean triple with b even.

Proof. We have $a^2 + b^2 = c^2$, so $b^2 = c^2 - a^2$ and hence

$$b^2 = (c-a)(c+a)$$

We claim that $\gcd(c-a, c+a) = 2$. Suppose that d is a common divisor of $c-a$ and $c+a$. We then have that $d \mid [(c-a)+(c+a)]$, so $d \mid 2c$. We also have $d \mid [(c+a)-(c-a)]$, so $d \mid 2a$. It follows that $d \mid \gcd(2b, 2c)$. Since $\gcd(2b, 2c) = 2 \cdot \gcd(b,c) = 2 \cdot 1 = 2$, it follows that either $d = 1$ or $d = 2$. Since a and c are both odd, we conclude that $c-a$ and $c+a$ are both even. Therefore, $\gcd(c-a, c+a) = 2$.

Fix $r, s, t \in \mathbb{N}^+$ with $c - a = 2r$, $c + a = 2s$, and $b = 2t$. We then have

$$(2t)^2 = 2r \cdot 2s$$

so

$$4t^2 = 4rs$$

and hence

$$t^2 = rs$$

Notice that $\gcd(r, s) = 1$ (if $d > 1$ is a common divisor of r and s, then $2d > 2$ is a common divisor of $c - a$ and $c + a$). Since rs is a square and $\gcd(r, s) = 1$, it follows that each of r and s are squares. Fix $m, n \in \mathbb{N}^+$ with $r = m^2$ and $s = n^2$. We now have

$$c - a = 2m^2 \qquad \text{and} \qquad c + a = 2n^2$$

Therefore

$$2a = (c + a) - (c - a) = 2n^2 - 2m^2 = 2(n^2 - m^2)$$

so $a = n^2 - m^2$. We also have

$$2c = (c - a) + (c + a) = 2m^2 - 2n^2 = 2(m^2 + n^2)$$

so $c = m^2 + n^2$. Finally, we have

$$b^2 = (c - a)(c + a) = 2m^2 \cdot 2n^2 = (2mn)^2$$

so $b = 2mn$. Notice that $m < n$ because $c - a < c + a$. Also, $\gcd(m, n) = 1$ because $\gcd(r, s) = 1$ and any common divisor of m and n is a common divisor of $r = m^2$ and $s = n^2$. Finally, m and n have distinct parities because otherwise each of a, b, c would be even, contrary to the fact that (a, b, c) is primitive.

We now prove the last statement. Let $m < n$ be relatively prime positive integers with distinct parities. Let

$$a = n^2 - m^2 \qquad b = 2mn \qquad c = m^2 + n^2$$

We then have

$$\begin{aligned} a^2 + b^2 &= (n^2 - m^2)^2 + (2mn)^2 \\ &= n^4 - 2m^2n^2 + m^4 + 4m^2n^2 \\ &= m^4 + 2m^2n^2 + n^4 \\ &= (m^2 + n^2)^2 \\ &= c^2 \end{aligned}$$

so (a, b, c) is a Pythagorean triple. Clearly, $b = 2mn$ is even. Thus, to finish the argument, we need only show that (a, b, c) is primitive. Suppose that $p \in \mathbb{Z}$ is a common divisor of a, b, and c. Notice that $a = n^2 - m^2$ is odd because m and n have distinct parities, so $p \neq 2$. Now p divides $c + a = 2n^2$ and also p divides $c - a = 2m^2$. Since p is an odd prime, this implies that p is a common divisor of m and n, which contradicts the fact that m and n are relaticely prime. $\square$

Chapter 2

Elementary Number Theory from an Algebraic Viewpoint

2.1 The Ring $\mathbb{Z}/n\mathbb{Z}$

Let $n \in \mathbb{N}^+$ and consider the principal ideal $n\mathbb{Z} = \langle n \rangle$. Since $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$, we may form the quotient ring $\mathbb{Z}n\mathbb{Z}$. Recall that elements of a quotient ring are additive cosets of $n\mathbb{Z}$. Thus, a typical element of $\mathbb{Z}/n\mathbb{Z}$ has the form $a + n\mathbb{Z}$ for some $a \in \mathbb{Z}$. Also recall that two cosets $a + n\mathbb{Z}$ and $b + n\mathbb{Z}$ are equal exactly when $a - b \in n\mathbb{Z}$, which is equivalent to saying that $n \mid a - b$. Thus, $a + n\mathbb{Z} = b + n\mathbb{Z}$ if and only if $a \equiv b \pmod{n}$. In other words, the relation on $\mathbb{Z}$ defined by equality of cosets is the same relation as modular arithmetic. In fact, the quotient ring construction is just a generalization of modular arithmetic. In what follows, we will typically write $\overline{a}$ rather than $a + n\mathbb{Z}$.

Recall that the key reason you studied ideals in abstract algebra was that the naive operations of addition and multiplication of cosets via representatives are well-defined. These naive definitions are

$$(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$$

and

$$(a + n\mathbb{Z}) \cdot (b + n\mathbb{Z}) = ab + n\mathbb{Z}$$

Saying that these operations are well-defined means that if $a + n\mathbb{Z} = c + n\mathbb{Z}$ and $b + n\mathbb{Z} = d + n\mathbb{Z}$, then

$$(a + b) + n\mathbb{Z} = (c + d) + \mathbb{Z}$$

and

$$ab + n\mathbb{Z} = cd + n\mathbb{Z}$$

Restated in terms of modular arithmetic, this says that if $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$, then

$$a + b \equiv c + d \pmod{n}$$

and

$$ab \equiv cd \pmod{n}$$

If you have forgotten how to justify these well-defined properties directly (rather than from the more general ring-theoretic construction), you should work through them.

To summarize, given $a, b \in \mathbb{Z}$, the following all express the same thing.

- $\overline{a} = \overline{b}$

- $a + n\mathbb{Z} = b + n\mathbb{Z}$
- $a \equiv b \pmod{n}$
- $n \mid (a - b)$
- $n \mid (b - a)$

Much of this chapter is devoted to the study of the commutative ring $\mathbb{Z}/n\mathbb{Z}$. As an additive group, $\mathbb{Z}/n\mathbb{Z}$ is fairly easy to understand because it is trivially a cyclic group of order n generated by $\overline{1}$. However, the multiplicative structure is much more interesting. In particular, we will spend time trying to understand the multiplicative group of units $U(\mathbb{Z}/n\mathbb{Z})$.

2.2 Euler's Theorem and Fermat's Theorem

Recall that for any $a \in \mathbb{Z}$, we have that $\overline{a} \in U(\mathbb{Z}/n\mathbb{Z})$ if and only if $\gcd(a, n) = 1$. Thus, $|U(\mathbb{Z}/n\mathbb{Z})|$ is the number of integers $a \in \{0, 1, 2, \ldots, n-1\}$ such that $\gcd(a, n) = 1$. We introduce a special function to describe this situation.

Definition 2.2.1. *We define a function $\varphi\colon \mathbb{N}^+ \to \mathbb{N}^+$ as follows. For each $n \in \mathbb{N}^+$, we let*

$$\varphi(n) = |\{a \in \{0, 1, 2, \ldots, n-1\} : \gcd(a, n) = 1\}|$$

The function φ is called the Euler φ-function or Euler totient function.

Therefore, by definition, we have $|U(\mathbb{Z}/n\mathbb{Z})| = \varphi(n)$ for all $n \in \mathbb{N}^+$. Notice that $\varphi(p) = p - 1$ for all primes p because all numbers in the set $\{1, 2, \ldots, p-1\}$ are relatively prime to p.

Theorem 2.2.2 (Euler's Theorem)**.** *Let $a \in \mathbb{Z}$ and $n \in \mathbb{N}^+$. If $\gcd(a, n) = 1$, then $a^{\varphi(n)} \equiv 1 \pmod{n}$.*

Algebraic Proof. Consider the group $G = U(\mathbb{Z}/n\mathbb{Z})$. Since $\gcd(a, n) = 1$, we have that $\overline{a}$ is a unit in $\mathbb{Z}/n\mathbb{Z}$ and so is an element of the multiplicative group G. Now $|G| = \varphi(n)$ from above, so by Lagrange's Theorem in group theory, we know that $|\overline{a}|$ divides $|G| = \varphi(n)$. Thus, $\overline{a}^{\varphi(n)} = \overline{1}$ in G, i.e. $a^{\varphi(n)} \equiv 1 \pmod{n}$. □

Elementary Proof. Let $b_1, b_2, \ldots, b_{\varphi(n)}$ be the numbers between 0 and $n-1$ which are relatively prime to n. Consider the list of numbers $ab_1, ab_2, \ldots, ab_{\varphi(n)}$. Notice that each of these numbers is relatively prime to n, and none of them are congruent to the others (because a has a multiplicative inverse modulo n). Thus, each element of the latter list of numbers is congruent to exactly one of the numbers in the former list. It follows that

$$b_1 \cdot b_2 \cdots b_{\varphi(n)} \equiv (ab_1) \cdot (ab_2) \cdots (ab_{\varphi(n)}) \pmod{n}$$

and thus

$$(b_1 \cdot b_2 \cdots b_{\varphi(n)}) \equiv a^{\varphi(n)} \cdot (b_1 \cdot b_2 \cdots b_{\varphi(n)}) \pmod{n}$$

Now the product $b_1 \cdot b_2 \cdots b_{\varphi(n)}$ is relatively prime to n because each b_i is relatively prime to n. Multiplying both sides by the inverse of this element, we conclude that

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

□

Corollary 2.2.3 (Fermat's Little Theorem)**.** *Suppose that $p \in \mathbb{N}^+$ is prime.*

- *If $a \in \mathbb{Z}$ and $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$.*

- *For all $a \in \mathbb{Z}$, we have $a^p \equiv a \pmod{p}$.*

Proof. Since p is prime, we know that $\varphi(p) = p-1$. Thus, the first part follows immediately from the Euler's Theorem. We now prove the second part. Suppose that $a \in \mathbb{Z}$ is arbitrary. If $p \mid a$, then $p \mid a^p$ trivially, so $a^p \equiv 0 \equiv a \pmod{p}$. On the other hand, if $p \nmid a$, then $\gcd(a,p) = 1$, so $a^{p-1} \equiv 1 \pmod{p}$ by the first part. Multiplying both sides of this by a, we conclude that $a^p \equiv a \pmod{p}$. □

2.3 Chinese Remainder Theorem

Theorem 2.3.1 (Chinese Remainder Theorem - Elementary Version with Two Moduli)**.** *Suppose that the numbers $m, n \in \mathbb{Z}$ are relatively prime and $a, b \in \mathbb{Z}$. There exists $x \in \mathbb{Z}$ such that*

$$x \equiv a \pmod{m} \qquad \textit{and} \qquad x \equiv b \pmod{n}$$

Furthermore, if $x_0 \in \mathbb{Z}$ is one solution to the above congruences, then an arbitrary $x \in \mathbb{Z}$ is also a solution if and only if $x \equiv x_0 \pmod{mn}$.

Proof 1 - Elementary Nonconstructive Proof. Let $m, n \in \mathbb{Z}$ be such that $\gcd(m,n) = 1$. Consider the mn many numbers in the set $S = \{0, 1, 2, \ldots, mn-1\}$. Now if $i, j \in S$ are such that $i \equiv j \pmod{m}$ and $i \equiv j \pmod{n}$, then $m \mid (i-j)$ and $n \mid (i-j)$, so $mn \mid (i-j)$ because $\gcd(m,n) = 1$, which implies that $i = j$ because $-mn < i-j < mn$. Therefore each of the mn many numbers in S given distinct pairs of remainders upon division by m and n. Since the number of such pairs is $m \cdot n$ and $|S| = mn$, it follows that every pair of remainders appears exactly once. In particular, given $a, b \in \mathbb{Z}$, there exists $x_0 \in S$ such that both

$$x_0 \equiv a \pmod{m} \qquad \text{and} \qquad x_0 \equiv b \pmod{n}$$

We now verify the last statement. Let $a, b \in \mathbb{Z}$. Suppose that x_0 is one solution. Suppose first that $x \equiv x_0 \pmod{mn}$. We then have that $x \equiv x_0 \equiv a \pmod{m}$ and $x \equiv x_0 \equiv b \pmod{n}$, so x is also a solution. Suppose conversely that $x \equiv a \pmod{m}$ and $x \equiv a \pmod{n}$. We then have that $x \equiv x_0 \pmod{m}$ and $x \equiv x_0 \pmod{n}$, so $m \mid (x - x_0)$ and $n \mid (x - x_0)$. Since m and n are relatively prime, it follows that $mn \mid (x - x_0)$, hence $x \equiv x_0 \pmod{mn}$. □

Proof 2 - Elementary Constructive Proof. Since $gcd(m,n) = 1$, we may fix $k, \ell \in \mathbb{Z}$ with $km + \ell n = 1$. Notice that $km \equiv 1 \pmod{n}$ so $\overline{k}$ is the multiplicative inverse of $\overline{m}$ in $\mathbb{Z}/n\mathbb{Z}$. Similarly, we have $\ell n \equiv 1 \pmod{m}$, so $\overline{\ell}$ is the multiplicative inverse of $\overline{m}$ in $\mathbb{Z}/m\mathbb{Z}$. Let $x_0 = bkm + a\ell n$. We check that x_0 satisfies the above congruences:

- Since $\ell n \equiv 1 \pmod{m}$, we have $a\ell n \equiv a \pmod{m}$. Now $bkm \equiv 0 \pmod{m}$, so adding these congruences we get $x_0 \equiv a \pmod{m}$.
- Since $km \equiv 1 \pmod{n}$, we have $bkm \equiv b \pmod{n}$. Now $a\ell n \equiv 0 \pmod{n}$, so adding these congruences we get $x_0 \equiv b \pmod{n}$.

The verification of the last statement is identical to the proof above. □

Example 2.3.2. *Find all $x \in \mathbb{Z}$ which simultaneously satisfy*

$$x \equiv 3 \pmod{14} \qquad \textit{and} \qquad x \equiv 8 \pmod{9}$$

Solution. Notice that $2 \cdot 14 + (-3) \cdot 9 = 1$ which can be found by inspection or by the Euclidean Algorithm:

$$\begin{aligned} 14 &= 1 \cdot 9 + 5 \\ 9 &= 1 \cdot 5 + 4 \\ 5 &= 1 \cdot 4 + 1 \\ 4 &= 4 \cdot 1 + 0 \end{aligned}$$

Therefore, working backwards, we have

$$\begin{aligned} 1 &= 5 - 4 \\ &= 5 - (9 - 5) \\ &= (-1) \cdot 9 + 2 \cdot 5 \\ &= (-1) \cdot 9 + 2 \cdot (14 - 9) \\ &= 2 \cdot 14 + (-3) \cdot 9 \end{aligned}$$

The proof of the Chinese Remainder Theorem lets $x_0 = 8 \cdot 2 \cdot 14 + 3 \cdot (-3) \cdot 9 = 143$. Thus, the complete solution is all x which satisfy $x \equiv 143 \pmod{126}$, i.e. all x which satisfy $x \equiv 17 \pmod{126}$. □

Although the elementary proofs are fairly direct and straightforward, we get much more information and elegance by abstracting the above ideas to the following version. There are no fundamentally new ideas in the this proof compared to the elementary one, but

Theorem 2.3.3 (Chinese Remainder Theorem - Algebraic Version with Two Moduli)**.** *If $m, n \in \mathbb{Z}$ are relatively prime, then*

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

as rings via the map $\phi(k + (mn)\mathbb{Z}) = (k + m\mathbb{Z}, k + n\mathbb{Z})$. In particular, the map ϕ is surjective so for all $a, b \in \mathbb{Z}$, there exists a unique $k \in \{0, 1, 2, \ldots, mn - 1\}$ such that

$$k + m\mathbb{Z} = a + m\mathbb{Z} \qquad \text{and} \qquad k + n\mathbb{Z} = b + n\mathbb{Z}$$

Proof. Define a function $\psi \colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ by $\phi(k) = (k + m\mathbb{Z}, k + n\mathbb{Z})$. We then have that ϕ is a ring homomorphism, and

$$\begin{aligned} ker(\psi) &= \{k \in \mathbb{Z} : \psi(k) = (0 + m\mathbb{Z}, 0 + n\mathbb{Z})\} \\ &= \{k \in \mathbb{Z} : (k + m\mathbb{Z}, k + n\mathbb{Z}) = (0 + m\mathbb{Z}, 0 + n\mathbb{Z})\} \\ &= \{k \in \mathbb{Z} : k + m\mathbb{Z} = 0 + m\mathbb{Z} \text{ and } k + n\mathbb{Z} = 0 + n\mathbb{Z}\} \\ &= \{k \in \mathbb{Z} : m \mid k \text{ and } n \mid k\} \\ &= \{k \in \mathbb{Z} : mn \mid k\} && (\text{since } gcd(m, n) = 1) \\ &= (mn)\mathbb{Z} \end{aligned}$$

Therefore, by the First Isomorphism Theorem, we have

$$\mathbb{Z}/(mn)\mathbb{Z} \cong ran(\psi)$$

as rings via the function

$$k + (mn)\mathbb{Z} \mapsto (k + m\mathbb{Z}, k + n\mathbb{Z})$$

Now both of rings $\mathbb{Z}/(mn)\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ have mn many elements, and the map above is injective, so it must be surjective. Alternatively, one can prove that the function is surjective constructively using the elementary proof. Thus

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

Both the existence and uniqueness statements follow. □

The fact that the above function is surjective is really just the elementary version of the Chinese Remainder Theorem. However, the algebraic proof shows that the function taking an integer modulo mn to the pair of remainders modulo m and n, gives a ring isomorphism. Thus, this "breaking up" of an integer modulo mn into the pair modulo m and modulo n separately preserves both addition and multiplication. This added information is very useful and will allow us to reduce certain problems to easier ones.

Example 2.3.4. *Find all $x \in \mathbb{Z}$ which simultaneously satisfy*

$$x^2 \equiv 4 \pmod 6 \qquad \textit{and} \qquad x^3 \equiv 3 \pmod 5$$

Solution. Simple inspection shows that:

- $x^2 \equiv 4 \pmod 6$ if and only if either $x \equiv 2 \pmod 6$ or $x \equiv 4 \pmod 6$.
- $x^3 \equiv 3 \pmod 5$ if and only if $x \equiv 2 \pmod 5$.

Therefore, an integer $x \in \mathbb{Z}$ satisfies our original two equations if and only if either:

- $x \equiv 2 \pmod 6$ and $x \equiv 2 \pmod 5$.
- $x \equiv 4 \pmod 6$ and $x \equiv 2 \pmod 5$.

The solution to the first (either by the following the constructive proof or by inspection) is $x \equiv 2 \pmod{30}$. The solution to the second is $x \equiv 22 \pmod{30}$. Thus, $x \in \mathbb{Z}$ is a solution to our original congruences if and only if either $x \equiv 2 \pmod{30}$ or $x \equiv 22 \pmod{30}$. □

Theorem 2.3.5 (Chinese Remainder Theorem - Abstract Version with Two Ideals)**.** *Let R be a commutative ring. Let I and J be ideals of R which are comaximal, i.e. $I + J = R$. We then have that $IJ = I \cap J$ and the map $\phi\colon R \to R/I \times R/J$ defined by*

$$\phi(r) = (r + I, r + J)$$

is a surjective homomorphism with kernel $IJ = I \cap J$. Therefore, by the First Isomorphism Theorem, we have

$$R/IJ \cong R/I \times R/J$$

Proof. Since $I + J = R$, we may fix $x \in I$ and $y \in J$ with $x + y = 1$. The map ϕ is clearly a ring homomorphism, but we need to check that $IJ = I \cap J$, that $ker(\phi) = I \cap J$, and that ϕ is surjective.

We always have $IJ \subseteq I \cap J$ for any ideals I and J. We need to show that $I \cap J \subseteq IJ$. For any $c \in I \cap J$, we have

$$c = c \cdot 1 = c \cdot (x + y) = xc + cy \in IJ$$

Therefore, $I \cap J \subseteq IJ$, and hence $I \cap J = IJ$.

We have

$$\begin{aligned} ker(\phi) &= \{a \in R : \phi(a) = (0 + I, 0 + J)\} \\ &= \{a \in R : (a + I, a + J) = (0 + I, 0 + J)\} \\ &= \{a \in R : a \in I \text{ and } a \in J\} \\ &= I \cap J \end{aligned}$$

Finally, we must check that ϕ is surjective. Fix $a, b \in R$. We need to find $r \in R$ with $\phi(r) = (a+I, b+I)$. In other words, we must find $r \in R$ such that $r + I = a + I$ and $r + J = b + J$, i.e. $r - a \in I$ and $r - b \in J$. Let $r = xb + ya$. We then have

$$\begin{aligned} r - a &= xb + ya - a \\ &= xb + (y - 1)a \\ &= xb + (-x)a \\ &= (b - a)x \end{aligned}$$

so $r - a \in I$. We also have

$$\begin{aligned} r - b &= xb + ya - b \\ &= (x-1)b + ya \\ &= (-y)b + ya \\ &= (a-b)y \end{aligned}$$

so $r - b \in J$. □

Theorem 2.3.6 (Chinese Remainder Theorem - Elementary Version)**.** *Suppose that $m_1, m_2, \ldots, m_\ell \in \mathbb{Z}$ are pairwise relatively prime and $a_1, a_2, \ldots, a_\ell \in \mathbb{Z}$. There exists $x \in \mathbb{Z}$ such that*

$$x \equiv a_i \pmod{m_i}$$

for all i. Furthermore, if $x_0 \in \mathbb{Z}$ is one solution to the above congruences, then an arbitrary $x \in \mathbb{Z}$ is also a solution if and only if $x \equiv x_0 \pmod{n}$.

Proof. Let $n = m_1 m_2 \cdots m_k$. For any i, the numbers $\frac{n}{m_i}$ and m_i are relatively prime, so we may fix ℓ_i such that $\frac{n}{m_i} \cdot \ell_i \equiv 1 \pmod{m_i}$. Let

$$x_0 = \sum_{i=1}^{k} \frac{n}{m_i} \cdot \ell_i a_i$$

We check that x_0 satisfies the above congruences.

Fix an arbitrary i. Since $\frac{n}{m_i} \cdot \ell_i \equiv 1 \pmod{m_i}$, we have $\frac{n}{m_i} \cdot \ell_i a_i \equiv a_i \pmod{m_i}$. Now for any $j \neq i$, we have that $m_i \mid \frac{n}{m_j}$, so $\frac{n}{m_j} \cdot \ell_j a_j \equiv 0 \pmod{m_i}$. Adding together each of the congruences, we conclude that $x_0 \equiv a_i \pmod{m_i}$.

We now verify the last statement. Suppose that x_0 is one solution. Suppose first that $x \equiv x_0 \pmod{n}$. We then have that $x \equiv x_0 \equiv a_i \pmod{m_i}$ for all i, so x is also a solution. Suppose conversely that $x \equiv a_i \pmod{m_i}$ for all i. We then have that $x \equiv x_0 \pmod{m_i}$ for all i, so $m_i \mid (x - x_0)$ for all i. Since the m_i are pairwise relatively prime, it follows that $n \mid (x - x_0)$, hence $x \equiv x_0 \pmod{n}$. □

Example 2.3.7. *Find all integers x such that $x^3 \equiv 53 \pmod{120}$.*

Solution. Notice that $120 = 12 \cdot 10 = 4 \cdot 3 \cdot 10 = 2^2 \cdot 3 \cdot 2 \cdot 5 = 2^3 \cdot 3 \cdot 5$. Notice that since 8, 3, and 5 are relatively prime in pairs, we have that $x^3 \equiv 53 \pmod{120}$ if and only if the following three congruences are true:

$$x^3 \equiv 53 \pmod 8 \qquad x^3 \equiv 53 \pmod 3 \qquad x^3 \equiv 53 \pmod 5$$

i.e. if and only if

$$x^3 \equiv 5 \pmod 8 \qquad x^3 \equiv 2 \pmod 3 \qquad x^3 \equiv 3 \pmod 5$$

By inspection in these small cases, these congruences are equivalent to

$$x \equiv 5 \pmod 8 \qquad x \equiv 2 \pmod 3 \qquad x \equiv 2 \pmod 5$$

We have $n = 8 \cdot 3 \cdot 5$. Our first goal is to find integers ℓ_1, ℓ_2, and ℓ_3 such that

$$15\ell_1 \equiv 1 \pmod 8 \qquad 40\ell_2 \equiv 1 \pmod 3 \qquad 24\ell_3 \equiv 1 \pmod 5$$

i.e. such that

$$7\ell_1 \equiv 1 \pmod 8 \qquad \ell_2 \equiv 1 \pmod 3 \qquad 4\ell_3 \equiv 1 \pmod 5$$

Thus, we may take $\ell_1 = 7$, $\ell_2 = 1$, and $\ell_3 = 4$. The proof of the Chinese Remainder Theorem lets

$$x_0 = 15 \cdot 7 \cdot 5 + 40 \cdot 1 \cdot 2 + 24 \cdot 4 \cdot 2 = 797$$

Thus, the complete solution is all x which satisfy $x \equiv 797 \pmod{120}$, i.e. all x which satisfy $x \equiv 77 \pmod{120}$. □

Theorem 2.3.8 (Chinese Remainder Theorem - Algebraic Version)**.** *If $m_1, m_2, \ldots, m_\ell \in \mathbb{Z}$ are pairwise relatively prime, then*

$$\mathbb{Z}/(m_1 m_2 \cdots m_\ell)\mathbb{Z} \cong \mathbb{Z}/m_2\mathbb{Z} \times \mathbb{Z}/m_2\mathbb{Z} \times \cdots \times \mathbb{Z}/m_\ell\mathbb{Z}$$

as rings via the map

$$\phi(k + m_1 m_2 \cdots m_\ell\mathbb{Z}) = (k + m_1\mathbb{Z}, k + m_2\mathbb{Z}, \ldots, k + m_\ell\mathbb{Z})$$

In particular, the map ϕ is surjective so for all $a_i \in \mathbb{Z}$, there exists a unique $k \in \{0, 1, 2, \ldots, m_1 m_2 \cdots m_\ell - 1\}$ such that

$$k + m_i\mathbb{Z} = a_i + m_i\mathbb{Z}$$

for all i.

Theorem 2.3.9 (Chinese Remainder Theorem - Abstract Version)**.** *Let R be a commutative ring. Let $I_1, I_2, \ldots, I_\ell$ be ideals of R which are comaximal in pairs, i.e. $I_i + I_j = R$ whenever $i \neq j$. We then have that $I_1 I_2 \cdots I_\ell = I_1 \cap I_2 \cap \cdots \cap I_\ell$ and the map $\phi\colon R \to R/I_1 \times R/I_2 \times \cdots \times R/I_\ell$ defined by*

$$\phi(r) = (r + I_1, r + I_2, \ldots, r + I_\ell)$$

is a surjective homomorphism with kernel $I_1 I_2 \cdots I_\ell = I_1 \cap I_2 \cap \cdots \cap I_\ell$. Therefore, by the First Isomorphism Theorem, we have

$$R/I_1 I_2 \cdots I_\ell \cong R/I_1 \times R/I_2 \times \ldots R/I_\ell$$

2.4 The Euler Function

Proposition 2.4.1. *Let R and S be commutative rings. We then have that $U(R \times S)$ are precisely the elements of the form (u, w) where $u \in U(R)$ and $w \in U(S)$. In other words, $U(R \times S) = U(R) \times U(S)$.*

Proof. Suppose first that $(u, w) \in U(R \times S)$. We may then fix $(x, y) \in R \times S$ with $(u, w) \cdot (x, y) = (1_R, 1_S)$. This implies that $(ux, wy) = (1_R, 1_S)$, so $ux = 1_R$ and $wy = 1_S$. Thus, $u \in U(R)$ and $w \in U(S)$, so $(u, w) \in U(R) \times U(S)$.

Suppose conversely that $(u, w) \in U(R) \times U(S)$. We then have that $u \in U(R)$ and $w \in U(S)$, so we may fix $x \in R$ and $y \in S$ with $ux = 1_R$ and $wy = 1_S$. We then have

$$(u, w) \cdot (x, y) = (ux, wy) = (1_R, 1_S)$$

which is the multiplicative identity of $R \times S$, so $(u, w) \in U(R \times S)$. □

Proposition 2.4.2. *Suppose that R and S are commutative rings and $\psi\colon R \to S$ is a ring isomorphism. We then have that $u \in U(R)$ if and only if $\psi(u) \in U(S)$. Furthermore, the function $\psi|_{U(R)}$ (that is, ψ restricted to $U(R)$) is an isomorphism of abelian groups from $U(R)$ onto $U(S)$.*

Proof. The first statement is immediate from the general idea that isomorphisms preserve all algebraic properties. However, here is a formal proof. Suppose that $u \in U(R)$. We may then fix $w \in R$ with $uw = 1_R$. Applying ψ we conclude that $\psi(uw) = \psi(1_R) = 1_S$, so $\psi(u) \cdot \psi(w) = 1_S$. Thus, $\psi(u) \in U(S)$. Suppose conversely that $u \in R$ and $\psi(u) \in U(S)$. Fix $z \in S$ with $\psi(u) \cdot z = 1_S$. Since ψ is surjective, we may fix $w \in R$ with $\psi(w) = z$. We then have $\psi(u) \cdot \psi(w) = 1_S$, so $\psi(uw) = \psi(1_R)$. Since ψ is injective, we conclude that $uw = 1$, so $u \in U(R)$.

We have shown that $u \in U(R)$ if and only if $\psi(u) \in U(S)$. Since ψ is surjective, we conclude that $\psi|_{U(R)}$ maps $U(R)$ bijectively onto $U(S)$. Using the fact that ψ is a ring homomorphism, it follows immediately that $\psi|_{U(R)} \colon U(R) \to U(S)$ preserves multiplication. Therefore, $\psi|_{U(R)}$ is an isomorphism of abelian groups from $U(R)$ onto $U(S)$. □

Corollary 2.4.3. *If $m, n \in \mathbb{Z}$ are relatively prime, then*

$$U(\mathbb{Z}/(mn)\mathbb{Z}) \cong U(\mathbb{Z}/m\mathbb{Z}) \times U(\mathbb{Z}/n\mathbb{Z})$$

as (multiplicative) abelian groups.

Proof. This follows immediately from the Chinese Remainder Theorem and the previous Proposition. □

Corollary 2.4.4. *If $m, n \in \mathbb{Z}$ are relatively prime, then $\varphi(mn) = \varphi(m) \cdot \varphi(n)$.*

Proof. If $m, n \in \mathbb{Z}$ are relatively prime, then

$$\begin{aligned} \varphi(mn) &= |U(\mathbb{Z}/(mn)\mathbb{Z})| \\ &= |U(\mathbb{Z}/m\mathbb{Z}) \times U(\mathbb{Z}/n\mathbb{Z})| \\ &= \varphi(m) \cdot \varphi(n) \end{aligned}$$

□

Proposition 2.4.5. *If $p \in \mathbb{N}^+$ is prime and $k \in \mathbb{N}^+$, then $\varphi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1)$.*

Proof. Fix a prime $p \in \mathbb{N}^+$ and $k \in \mathbb{N}^+$. We need to count the number of $m \in \mathbb{N}$ with $0 \leq m \leq p^k - 1$ such that $\gcd(m, p^k) = 1$. Instead, we count the complement, i.e. the number of $m \in \mathbb{N}$ with $0 \leq m \leq p^k - 1$ such that $\gcd(m, p^k) > 1$. Notice that the only positive divisors of p^k are the numbers in the set $\{1, p, p^2, \ldots, p^k\}$, so a number m fails to be relatively prime to p^k if and only if $p \mid m$. Thus, we count the numbers m with $0 \leq m \leq p^k - 1$ such that $p \mid m$. These numbers are

$$0p, 1p, 2p, 3p, \ldots, (p^{k-1} - 2)p, (p^{k-1} - 1)p$$

Thus, since we start counting with 0, there are p^{k-1} many numbers m with $0 \leq m \leq p^k - 1$ such that $\gcd(m, p^k) > 1$. Since we counted the complement, and there are p^k total many elements m with $0 \leq m \leq p^k - 1$, it follows that $\varphi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1)$. □

Corollary 2.4.6. *Let $n \in \mathbb{N}^+$ with $n \geq 2$. Write the prime factorization of n as*

$$n = p_1^{k_1} p_2^{k_2} \cdots p_\ell^{k_\ell}$$

where the p_i are distinct and each $k_i \in \mathbb{N}^+$. We then have

$$\begin{aligned} \varphi(n) &= \prod_{i=1}^{\ell} (p_i^{k_i} - p_i^{k_i - 1}) \\ &= \prod_{i=1}^{\ell} p_i^{k_i - 1}(p_i - 1) \\ &= n \cdot \prod_{i=1}^{\ell} \left(1 - \frac{1}{p_i}\right) \end{aligned}$$

Proof. This follows immediately from the two previous results. □

Before moving on to another important result about φ, we recall a few results from group theory.

Proposition 2.4.7. *Let G be a group and let $a \in G$. Suppose that $|a| = n \in \mathbb{N}^+$. For any $k \in \mathbb{Z}$, we have $a^k = e$ if and only if $n \mid k$.*

Proof. Let $|a| = n \in \mathbb{Z}$. We then have in particular that $a^n = e$. Suppose first that $k \in \mathbb{Z}$ is such that $n \mid k$. Fix $m \in \mathbb{Z}$ with $k = nm$ We then have

$$a^k = a^{nm} = (a^n)^m = e^m = e$$

so $a^k = e$. Suppose conversely that $k \in \mathbb{Z}$ and that $a^k = e$. Since $n > 0$, we may write $k = qn + r$ where $0 \leq r < n$. We then have

$$\begin{aligned} e &= a^k \\ &= a^{qn+r} \\ &= a^{qn}a^r \\ &= (a^n)^q a^r \\ &= e^q a^r \\ &= a^r \end{aligned}$$

Now by definition we know that n is the least positive power of a which gives the identity. Therefore, since $0 \leq r < n$ and $a^r = e$, we must have that $r = 0$. It follows that $k = qn$ so $n \mid k$. □

Proposition 2.4.8. *Let G be a group and let $a \in G$. Suppose that $|a| = n$. For any $k \in \mathbb{Z}$, we have*

$$|a^k| = \frac{n}{\gcd(n, k)}$$

Proof. Fix $k \in \mathbb{Z}$ and let $d = \gcd(n, k)$. The order of a^k is the least $m > 0$ such that $(a^k)^m = e$, i.e. the least m such that $n \mid km$. Notice that $\frac{n}{d}$ certainly works as such an m because

$$k \cdot \frac{n}{d} = n \cdot \frac{k}{d}$$

and $\frac{k}{d} \in \mathbb{Z}$ because $d \mid k$. We now need to show that $\frac{n}{d}$ is the least positive value of m that works. Suppose then that $m > 0$ and $n \mid km$. Fix $\ell \in \mathbb{Z}$ such that $n\ell = km$. Dividing through by d gives

$$\frac{n}{d} \cdot \ell = \frac{k}{d} \cdot m$$

hence

$$\frac{n}{d} \mid \frac{k}{d} \cdot m$$

Since $\gcd(\frac{n}{d}, \frac{k}{d}) = 1$ (this is easy to verify because $d = \gcd(n, k)$), it follows that $\frac{n}{d} \mid m$. Since $\frac{n}{d}$ and m are each positive, we conclude that $\frac{n}{d} \leq m$. Therefore, $\frac{n}{d}$ is the least m such that $n \mid km$, and so we conclude that $|a^k| = \frac{n}{d}$. □

Corollary 2.4.9. *If G is a cyclic group of order n, then G has exactly $\varphi(n)$ many elements of order n.*

Proof. Since G is cyclic, we may fix a generator $c \in G$, i.e. an element with $|c| = n$. We then have that $G = \{c^0, c^1, c^2, \ldots, c^{n-1}\}$ and furthermore if $0 \leq i < j < n$, then $c^i \neq c^j$. The previous result implies that given $k \in \{0, 1, 2, \ldots, n-1\}$, we have $|c^k| = n$ if and only if $\gcd(k, n) = 1$. Therefore, the number of elements of G of order n equals $\varphi(n)$. □

Proposition 2.4.10. *Let G be a cyclic group of order n and let $d \mid n$. We then have that G has exactly $\varphi(d)$ many elements of order d.*

Proof. Fix a generator $c \in G$. Notice that $|c^{n/d}| = d$ (either directly or from the above result), so G has an element of order d. Let $H = \langle c^{n/d} \rangle$ and notice that H is a cyclic group of order d. Therefore, H has exactly $\varphi(d)$ many elements of order d from above. To complete the proof, we need only show that every element of G with order d must be an element of H.

Suppose then that $g \in G$ is an element with $|g| = d$. Since c is a generator of G, we may fix $k \in \mathbb{Z}$ with $g = c^k$. Since $|c^k| = |g| = d$, we must have that

$$\frac{n}{\gcd(k,n)} = d$$

and thus $\gcd(k,n) = \frac{n}{d}$. It follows that $\frac{n}{d} \mid k$ and hence $g = c^k \in \langle c^{n/d} \rangle = H$. Therefore, every element of order d is an element of H. This completes the proof. □

Theorem 2.4.11. *For any $n \in \mathbb{N}^+$, we have*

$$n = \sum_{d \mid n} \varphi(d)$$

where the summation is over all positive divisors d of n.

Proof. Let $n \in \mathbb{N}^+$. Fix any cyclic group of order G. We know that every element of G has order some divisor of n, and furthermore we know that if $d \mid n$, then G has exactly $\varphi(d)$ many elements of order d. Therefore, the sum on the right-hand side simply counts the number of elements of G by breaking them up into the various possible orders. Since $|G| = n$, the result follows. □

Notice that the above proposition gives a recursive way to calculate $\varphi(n)$ because

$$\varphi(n) = n - \sum_{\substack{d \mid n \\ d < n}} \varphi(d)$$

2.5 Wilson's Theorem

Proposition 2.5.1. *Let G be a finite abelian group with elements $a_1, a_2, \cdots, a_n$. Let $b_1, b_2, \ldots, b_k$ be the elements of G which satisfy $x^2 = e$ (i.e. the elements which are their own inverses). We then have*

$$\prod_{i=1}^{n} a_i = \prod_{i=1}^{k} b_i$$

Proof. Let $a_i \in G$ and suppose that $a_i^{-1} = a_j$ for some $j \neq i$. In the product of all elements of G, we can pair off a_i with a_j and have these elements cancel each other (notice that $a_j^{-1} = a_i$ so a_j gets paired with a_i as well). Thus, all elements which are not their own inverses disappear from the product, and the product simplifies into the product of only those elements which are their own inverses. □

Proposition 2.5.2. *Let p be prime. The elements of $U(\mathbb{Z}/p\mathbb{Z})$ which are their own inverse are exactly $\overline{1}$ and $\overline{-1} = \overline{p-1}$. Furthermore, if p is an odd prime, then $\overline{1} \neq \overline{-1}$.*

Proof. First notice that $\overline{1} \cdot \overline{1} = \overline{1 \cdot 1} = \overline{1}$ and $\overline{-1} \cdot \overline{-1} = \overline{(-1) \cdot (-1)} = \overline{1}$, so both of these elements are their own inverses. Furthermore, if $\overline{1} = \overline{-1}$, then $p \mid 2$, so $p = 2$.

Suppose now that $\overline{k} \in U(\mathbb{Z}/p\mathbb{Z})$ is its own inverse. We then have that $k^2 \equiv 1 \pmod{p}$, so $p \mid (k^2 - 1)$ and thus $p \mid (k-1)(k+1)$. Since p is prime, this implies that either $p \mid (k-1)$ or $p \mid (k+1)$, i.e. either $k \equiv 1 \pmod{p}$ or $k \equiv -1 \pmod{p}$. Thus, either $\overline{k} = \overline{1}$ or $\overline{k} = \overline{-1}$. □

Theorem 2.5.3 (Wilson's Theorem)**.** *If p is prime, then $(p-1)! \equiv -1 \pmod{p}$.*

Proof 1. Notice that in the ring $\mathbb{Z}/p\mathbb{Z}$, we have

$$\overline{(p-1)!} = \overline{1 \cdot 2 \cdot 3 \cdots (p-1)} = \overline{1} \cdot \overline{2} \cdot \overline{3} \cdots (\overline{p-1})$$

Thus, $\overline{(p-1)!}$ is the product of all the elements in the abelian group $U(\mathbb{Z}/p\mathbb{Z})$. By the above two propositions, this equals

$$\overline{1} \cdot \overline{p-1} = \overline{1} \cdot \overline{-1} = \overline{-1}$$

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

Proof 2. The result is trivial if $p = 2$, so assume that p is odd. Work over the field $F = \mathbb{Z}/p\mathbb{Z}$. Consider the polynomial $f(x) = x^{p-1} - \overline{1}$ in $F[x]$. Notice that every element of $F - \{0\}$ is a root of this polynomial by Fermat's Little Theorem. Thus, for each $a \in F - \{0\}$, the irreducible polynomial $x - a$ divides $x^{p-1} - \overline{1}$ in $F[x]$. Now each polynomial of the form $x - a$ is irreducible in $F[x]$, so as $F[x]$ is UFD, we conclude that some associate of it must appear in any factorization of $x^{p-1} - \overline{1}$ into irreducibles. Therefore, we can write

$$x^{p-1} - \overline{1} = (x - \overline{1})(x - \overline{2}) \cdots (x - (\overline{p-1})) \cdot f(x)$$

for some $f(x) \in F[x]$. Comparing degrees on the left and right, we must have $\deg(f(x)) = 0$, so $f(x)$ is a constant. Comparing leading terms on both sides, it follows that $f(x) = \overline{1}$. Therefore, we have

$$x^{p-1} - \overline{1} = (x - \overline{1})(x - \overline{2}) \cdots (x - (\overline{p-1}))$$

By either plugging in $\overline{0}$ or examining the constant terms, we conclude that

$$\begin{aligned}
\overline{-1} &= \overline{-1} \cdot \overline{-2} \cdot \overline{-3} \cdots (\overline{-(p-1)}) \\
&= \overline{(-1)^{p-1}} \cdot \overline{1} \cdot \overline{2} \cdot \overline{3} \cdots (\overline{p-1}) \\
&= \overline{(-1)^{p-1}} \cdot \overline{(p-1)!} \\
&= \overline{(p-1)!} && \text{(since } p \text{ is odd)}
\end{aligned}$$

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

2.6 $U(\mathbb{Z}/p\mathbb{Z})$ is Cyclic

One of primary goals at this point is to understand the structure of the multiplicative group $U(\mathbb{Z}/n\mathbb{Z})$. Using the Chinese Remainder Theorem, it suffices to understand the structure of $U(\mathbb{Z}/p^k\mathbb{Z})$ for primes p. An example to keep in mind throughout this section is $U(\mathbb{Z}/8\mathbb{Z}) = \{\overline{1}, \overline{3}, \overline{5}, \overline{7}\}$. A simple calculation shows that $a^2 = 1$ for all $a \in U(\mathbb{Z}/8\mathbb{Z})$. Since this group has order 4, it is not cyclic. In fact, it turns out that $U(\mathbb{Z}/8\mathbb{Z})$ is isomorphic to the direct product of two copies of the cyclic group of order 2.

As we will see, for odd primes p, the group $U(\mathbb{Z}/p^k\mathbb{Z})$ is cyclic. In this section, we prove the result when $k = 1$, i.e. we prove that $U(\mathbb{Z}/p\mathbb{Z})$ is cyclic for every prime p. In fact, we will prove the following stronger result.

Theorem 2.6.1. *Suppose that F is a field, and suppose that G is a finite subgroup of the multiplicative group $U(F)$. We then have that G is cyclic.*

From this theorem, we obtain the corollary that is important to us.

Corollary 2.6.2. *The group $U(\mathbb{Z}/p\mathbb{Z})$ is cyclic for every prime p.*

Proof. Immediate from the above theorem because $\mathbb{Z}/p\mathbb{Z}$ is a field when p is prime. □

We give two proofs of this fact. The first uses properties of φ established above, while the second is more directly algebraic. The fundamental idea fueling both proofs is the fact that a polynomial of degree n has at most n roots when working over a field. Applying this to polynomials of the form $x^d - 1$ allows one to conclude that a finite subgroup G of $U(F)$ can not have too many elements of small order. The proofs then use this idea to argue for the existence of an element of order n.

Proof 1 of Theorem 2.6.1. Let $n = |G|$. For each positive $d \mid n$, let $f(d)$ be the number of elements of G of order d. Fix a positive $d \mid n$, and suppose that $f(d) \neq 0$. We may then fix an element $g \in G$ of order d. Let $H = \langle g \rangle$ and notice that $|H| = d$. By Lagrange's Theorem, we have $h^d = 1$ for all $h \in H$, so every element of H is a root of the polynomial $x^d - 1 \in F[x]$. Now F is a field, so we know that $x^d - 1$ has at most d roots in F, and so we have found all of them. Now every element of order d in G is a root of this polynomial, so every element of G of order d must be in H. From above, we know that H has $\varphi(d)$ many elements of order d. Thus, for all positive $d \mid n$, we know that either $f(d) = 0$ or $f(d) = \varphi(d)$. It follows that $f(d) \leq \varphi(d)$ for all positive $d \mid n$.

We know use this result to finish the proof. Since every element of G has order some positive divisor of n by Lagrange's Theorem, we have

$$\begin{aligned} n &= \sum_{d|n} f(d) \\ &\leq \sum_{d|n} \varphi(d) && \text{(from above)} \\ &= n && \text{(from above)} \end{aligned}$$

This implies that we must have $f(d) = \varphi(d)$ for all positive $d \mid n$. In particular, $f(n) = \varphi(n) > 0$, so G has an element of order n. Therefore, G is cyclic. □

Lemma 2.6.3. *Let G be a finite abelian group and let $g, h \in G$ with $|g| = m$, $|h| = n$, and $\gcd(m, n) = 1$. We then have that $|gh| = mn$.*

Proof. We have

$$\begin{aligned} (gh)^{mn} &= g^{mn}h^{mn} && \text{(since } G \text{ is abelian)} \\ &= (g^m)^n(h^n)^m \\ &= e^n e^m \\ &= e \end{aligned}$$

so $|gh| \leq mn$. Suppose that $k \in \mathbb{N}^+$ is such that $(gh)^k = e$. We then have $g^k h^k = e$. Raising each side to the m^{th} power gives $h^{km} = e$, and hence $n \mid km$. Since $\gcd(m, n) = 1$, it follows that $n \mid k$. Similarly, if we raise each side to the n^{th} power we get $g^{kn} = e$, so $m \mid kn$. Since $\gcd(m, n) = 1$, it follows that $m \mid k$. Since $m \mid k$ and $n \mid k$, and $\gcd(m, n) = 1$, it follows that $mn \mid k$, and hence $k \geq mn$. Therefore, $|gh| = mn$. □

Lemma 2.6.4. *Let G be a finite abelian group. Let ℓ be the smallest number such that $a^\ell = e$ for all $a \in G$ (i.e. ℓ is the least common multiple of the orders of all the elements of G). Then G has an element of order ℓ.*

Proof. Write the prime factorization of ℓ as

$$\ell = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$$

where the p_i are distinct primes and each $\alpha_i \geq 1$. Since ℓ is the least common multiple of the orders of all of the elements of G, it must be the case that $p_i^{\alpha_i}$ divides the order of some element of G for each i. Thus, for each i, we may fix an element $g_i \in G$ such that $|g_i| = p_i^{\alpha_i}$. Let

$$b = g_1 g_2 \cdots g_k$$

Using the lemma, we conclude that $|b| = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k} = \ell$. □

With these lemmas in hand, we are not ready to give our second proof of the above theorem.

Proof 2 of Theorem 2.6.1. Suppose that $|G| = n$. Let ℓ be the least common multiple of the orders of all of the elements of G. Notice that for all $a \in G$, we have that $|a|$ divides n be Lagrange's Theorem. Therefore, n is a common multiple of the orders of all the elements of G, so $\ell \leq n$. We also have that $x^\ell - 1$ has n roots in G, so we must have that $n \leq \ell$. Putting these together, we conclude that $\ell = n$. By the lemma, G has an element of order $\ell = n$, so G is cyclic. □

Definition 2.6.5. *Let $n \in \mathbb{N}^+$. An integer $a \in \mathbb{Z}$ is called a* primitive root *modulo n if $\overline{a} \in \mathbb{Z}/n\mathbb{Z}$ generates $U(\mathbb{Z}/n\mathbb{Z})$.*

Corollary 2.6.6. *For every prime p, the group $U(\mathbb{Z}/p\mathbb{Z})$ is cyclic, so there exists a primitive root modulo p.*

Example 2.6.7. *The number 2 is primitive root modulo 3, 5, 11, and 13, but not modulo 7. The number 3 is a primitive root modulo 7.*

2.7 Prime Powers

Before jumping into the general theory, we prove two important lemmas.

Lemma 2.7.1. *Let p be prime and let $i \in \mathbb{N}$ with $1 \leq i \leq p-1$. We then have that $p \mid \binom{p}{i}$.*

Proof. By definition, we have

$$\binom{p}{i} = \frac{p!}{i! \cdot (p-i)!}$$

so

$$i! \cdot (p-1)! \cdot \binom{p}{i} = p!$$

Notice that $ord_p(i!) = 0 = ord_p((p-1)!)$ since p does not divide any positive natural number less than p. Also, we have $ord_p(p!) = 1$. Applying ord_p to both sides of the above equation then gives

$$ord_p\left(\binom{p}{i}\right) = 1$$

Therefore, $p \mid \binom{p}{i}$. □

Lemma 2.7.2. *Suppose that p is prime and that $k \in \mathbb{N}^+$. If*

$$a \equiv b \pmod{p^k}$$

then

$$a^p \equiv b^p \pmod{p^{k+1}}$$

Proof. We have $p^k \mid (a-b)$, so we many fix $m \in \mathbb{Z}$ with $mp^k = a-b$. We then have that $a = b+mp^k$, hence

$$\begin{aligned} a^p &= (b+mp^k)^p \\ &= b^p + \binom{p}{1} b^{p-1}(mp^k)^1 + \sum_{i=2}^{p} \binom{p}{i} b^{p-i}(mp^k)^i \\ &= b^p + b^{p-1}mp^{k+1} + \sum_{i=2}^{p} \binom{p}{i} b^{p-i} m^i p^{ik} \\ &= b^p + b^{p-1}mp^{k+1} + p^{2k} \cdot \sum_{i=2}^{p} \binom{p}{i} b^{p-i} m^i p^{(i-2)k} \end{aligned}$$

Since $2k \geq k+1$ (as $k \geq 1$), it follows that $a^p - b^p$ is divisible by p^{k+1}, i.e. that $a^p \equiv b^p \pmod{p^{k+1}}$. □

2.7.1 Powers of 2

Proposition 2.7.3. *Suppose that $k \geq 3$. For all $a \in \mathbb{Z}$ with a odd, we have*

$$a^{2^{k-2}} \equiv 1 \pmod{2^k}$$

Proof. We prove the result by induction on k. For $k = 3$, we have $2^3 = 8$, and a straightforward calculation shows that each of $\overline{1}$, $\overline{3}$, $\overline{5}$, and $\overline{7}$ satisfy $a^2 = 1$ in $U(\mathbb{Z}/8\mathbb{Z})$.

Suppose that the result is true for a fixed $k \geq 3$. We prove it for $k+1$. Let $a \in \mathbb{Z}$ with a odd. By induction, we have

$$a^{2^{k-2}} \equiv 1 \pmod{2^k}$$

By Lemma 2.7.2, we conclude that

$$(a^{2^{k-2}})^2 \equiv 1^2 \pmod{2^{k+1}}$$

which says that

$$a^{2^{k-1}} \equiv 1 \pmod{2^{k+1}}$$

The result follows by induction. □

Corollary 2.7.4. *For every $k \geq 3$, the group $U(\mathbb{Z}/2^k\mathbb{Z})$ is not cyclic (i.e. there is no primitive root modulo 2^k whenever $k \geq 3$).*

Proof. Fix $k \geq 3$. We have that

$$|U(\mathbb{Z}/2^k\mathbb{Z})| = \varphi(2^k) = 2^{k-1}$$

Since a number is relatively prime to 2^k exactly when it is odd, the previous proposition tells us that every element of the group $U(\mathbb{Z}/2^k\mathbb{Z})$ has order at most 2^{k-2}. The result follows. □

Proposition 2.7.5. *For every $k \geq 3$, the element $\overline{5} \in U(\mathbb{Z}/2^k\mathbb{Z})$ has order 2^{k-2}.*

Proof. We know that $|U(\mathbb{Z}/2^k\mathbb{Z})| = \varphi(2^k) = 2^{k-1}$. Since $U(\mathbb{Z}/2^k\mathbb{Z})$ is not cyclic, we know that the order of $\overline{5}$ is not 2^{k-1}. Since the order of $\overline{5}$ must divide 2^{k-1} by Lagrange's Theorem, it must equal 2^ℓ for some $\ell \leq k-2$. Now $2^\ell \mid 2^{k-3}$ whenever $\ell \leq k-3$, so to show that the order of $\overline{5}$ is 2^{k-2} it suffices to show that

$$5^{2^{k-3}} \not\equiv 1 \pmod{2^k}$$

We prove the stronger statement that

$$5^{2^{k-3}} \equiv 2^{k-1} + 1 \pmod{2^k}$$

by induction on $k \geq 3$. When $k = 3$, this statement reads

$$5 \equiv 4 + 1 \pmod 8$$

which is trivially true. Suppose that the result is true for some fixed $k \geq 3$. We prove it for $k + 1$. The inductive hypothesis tells us that

$$5^{2^{k-3}} \equiv 2^{k-1} + 1 \pmod{2^k}$$

By Lemma 2.7.2, we conclude that

$$(5^{2^{k-3}})^2 \equiv (2^{k-1} + 1)^2 \pmod{2^{k+1}}$$

which says that

$$5^{2^{k-2}} \equiv 2^{2k-2} + 2 \cdot 2^{k-1} + 1 \pmod{2^{k+1}}$$

Now $2k - 2 \geq k + 1$ because $k \geq 3$, so $2^{2k-2} \equiv 0 \pmod{2^{k+1}}$. Since $2 \cdot 2^{k-1} = 2^k$, we see that

$$5^{2^{k-2}} \equiv 2^k + 1 \pmod{2^{k+1}}$$

The result follows by induction. □

Proposition 2.7.6. *Suppose that $k \geq 3$. In the group $U(\mathbb{Z}/2^k\mathbb{Z})$, we have $\overline{-1} \notin \langle \overline{5} \rangle$.*

Proof. We have $5 \equiv 1 \pmod 4$, and hence $5^\ell \equiv 1 \pmod 4$ for all $\ell \geq 1$. Therefore, $5^\ell \not\equiv -1 \pmod 4$ for each $\ell \geq 1$. Since $4 \mid 2^k$ (as $k \geq 3$), it follows that $5^\ell \not\equiv -1 \pmod{2^k}$ for each $\ell \geq 1$. Therefore, $\overline{-1} \notin \langle \overline{5} \rangle$. □

Corollary 2.7.7. *The group $U(\mathbb{Z}/2\mathbb{Z})$ is the trivial group, the group $U(\mathbb{Z}/4\mathbb{Z})$ is cyclic of order 2, and for all $k \geq 3$, the group $U(\mathbb{Z}/2^k\mathbb{Z})$ is the direct product of a cyclic group of order 2^{k-2} and a cyclic group of order 2.*

Proof. The statements for $U(\mathbb{Z}/2\mathbb{Z})$ and $U(\mathbb{Z}/4\mathbb{Z})$ are trivial. Fix $k \geq 3$ and consider the group $U(\mathbb{Z}/2^k\mathbb{Z})$. Let $H = \langle \overline{5} \rangle$ and let $K = \langle \overline{-1} \rangle$. We then have that $|H| = 2^{k-2}$, $|K| = 2$, and $H \cap K = \{\overline{1}\}$. Now HK is a subgroup of $U(\mathbb{Z}/2^k\mathbb{Z})$ because they are normal (as the group is abelian), so since H is a proper subgroup of HK and $|U(\mathbb{Z}/2^k\mathbb{Z})| = 2^{k-1} = 2 \cdot |H|$, we conclude that $HK = U(\mathbb{Z}/2^k\mathbb{Z})$. It follows that G is internal direct product of H and K, and therefore $U(\mathbb{Z}/2^k\mathbb{Z}) \cong H \times K$. □

2.7.2 Powers of Odd Primes

Fix an odd prime p. We first explore $U(\mathbb{Z}/p^2\mathbb{Z})$ in the hopes of finding a primitive root. Fix a primitive root g modulo p. Now we want to determine whether g is a primitive root modulo p^2. Let n be the order of $\overline{g}$ viewed as an element of $\mathbb{Z}/p^2\mathbb{Z}$. We then have that

$$g^n \equiv 1 \pmod{p^2}$$

so since $p \mid p^2$ we know that

$$g^n \equiv 1 \pmod p$$

Now the order of $\overline{g}$ when viewed as an element of $\mathbb{Z}/p\mathbb{Z}$ equals $p - 1$, so we must have that $p - 1 \mid n$. We also know that n divides $\varphi(p^2) = p(p - 1)$ by Lagrange's Theorem, so the only possibilities are that either $n = p - 1$ or $n = p(p - 1)$. Of course, if $n = p(p - 1)$, then we are happy because g will also be a primitive root modulo p^2. The problem is that this is not always guaranteed. For example, -1 is a primitive modulo 3 but is not a primitive root modulo 9. Also, 7 is a primitive root modulo 5 but is not a primitive root modulo 25 (since $7^4 \equiv 1 \pmod{25}$). For a more interesting example, 14 is a primitive root modulo 29, but

$$14^{28} \equiv 1 \pmod{841}$$

so 14 is not a primitive root modulo $29^2 = 841$.

What do we do when we end up in the unfortunate case where $n = p - 1$? We then have that

$$g^{p-1} \equiv 1 \pmod{p^2}$$

Now $g + p \equiv g \pmod{p}$, so $g + p$ is also a primitive root modulo p. Notice that

$$\begin{aligned}(g+p)^{p-1} &= g^{p-1} + \binom{p-1}{1} g^{p-2}p + \sum_{i=2}^{p-1} \binom{p-1}{i} g^{p-1-i}p^i \\ &= g^{p-1} + g^{p-2}p(p-1) + \sum_{i=2}^{p-1} \binom{p-1}{i} g^{p-1-i}p^i\end{aligned}$$

and thus

$$\begin{aligned}(g+p)^{p-1} &\equiv g^{p-1} + g^{p-2}p(p-1) \pmod{p^2} \\ &\equiv 1 + g^{p-2}p(p-1) \pmod{p^2}\end{aligned}$$

Now $p \nmid g$ and $p \nmid (p-1)$, so as p is prime we conclude that $p \nmid g^{p-2}(p-1)$. It follows that $p^2 \nmid g^{p-2}p(p-1)$, and so

$$(g+p)^{p-1} \not\equiv 1 \pmod{p^2}$$

Therefore, $g + p$ is a primitive root modulo p^2.

Theorem 2.7.8. *If p is an odd prime, then $U(\mathbb{Z}/p^2\mathbb{Z})$ is cyclic.*

Proof. Fix a primitive root g modulo p. If $g^{p-1} \not\equiv 1 \pmod{p^2}$, then the above argument shows that g is a primitive root modulo p^2. Suppose then that $g^{p-1} \equiv 1 \pmod{p^2}$. In this case, the number $g + p$ is a primitive root modulo p, and from above we have

$$(g+p)^{p-1} \not\equiv 1 \pmod{p^2}$$

Therefore, by the above argument applied to $g+p$ instead of g itself, we see that g is a primitive root modulo p^2. □

We now aim to push this result to higher powers of an odd prime p. Fix an odd prime p and an integer $k \geq 3$. Suppose that g is a primitive root modulo p^2. We want to determine whether g is a primitive root modulo p^k. Let n be the order of $\overline{g}$ viewed as an element of $\mathbb{Z}/p^k\mathbb{Z}$. Now

$$|U(\mathbb{Z}/p^k\mathbb{Z})| = \varphi(p^k) = p^{k-1}(p-1)$$

so our hope is to show that $n = p^{k-1}(p-1)$. We have

$$g^n \equiv 1 \pmod{p^k}$$

so since $p \mid p^k$ we know that

$$g^n \equiv 1 \pmod{p}$$

Now the order of $\overline{g}$ when viewed as an element of $\mathbb{Z}/p\mathbb{Z}$ equals $p - 1$ (because a primitive root modulo p^2 is automatically a primitive root modulo p), so we must have that $p - 1 \mid n$. By Lagrange's Theorem, we also know that $n \mid \varphi(p^k) = p^{k-1}(p-1)$. Therefore, we must have $n = p^m(p-1)$ for some $m \in \mathbb{N}$ with $0 \leq m \leq k-1$. Our goal is to show that $m = k - 1$.

Since $g^{p-1} \equiv 1 \pmod{p}$, we may write $g^{p-1} = 1 + ap$ for some $a \in \mathbb{Z}$. Now if $p \mid a$, then we would have

$$g^{p-1} \equiv 1 \pmod{p^2}$$

contrary to the fact that g is a primitive root modulo p^2. It follows that $p \nmid a$. We now prove the following.

Proposition 2.7.9. *For all $\ell \in \mathbb{N}$, we have*

$$g^{p^\ell(p-1)} \equiv 1 + ap^{\ell+1} \pmod{p^{\ell+2}}$$

Proof. We prove the result by induction. If $\ell = 0$ this is trivial because $g^{p-1} = 1 + ap$. Suppose that we know the result for some fixed $\ell \in \mathbb{N}$, i.e. we know that

$$g^{p^\ell(p-1)} \equiv 1 + ap^{\ell+1} \pmod{p^{\ell+2}}$$

Using Lemma 2.7.2, we conclude that

$$g^{p^{\ell+1}(p-1)} \equiv (1 + ap^{\ell+1})^p \pmod{p^{\ell+3}}$$

By the Binomial Theorem we know that

$$\begin{aligned}(1 + ap^{\ell+1})^p &= 1 + \binom{p}{1}ap^{\ell+1} + (\sum_{i=2}^{p-1}\binom{p}{i}(ap^{\ell+1})^i) + \binom{p}{p}(ap^{\ell+1})^p \\ &= 1 + ap^{\ell+2} + (\sum_{i=2}^{p-1}\binom{p}{i}a^i p^{i(\ell+1)}) + a^p p^{p(\ell+1)}\end{aligned}$$

Now $p \geq 3$, so

$$p(\ell+1) \geq 3(\ell+1) = 3\ell + 3 \geq \ell + 3$$

so $p^{\ell+3}$ divides the last term in the above sum. For any i with $2 \leq i \leq p-1$, we have

$$i(\ell+1) \geq 2(\ell+1) = 2\ell + 2 \geq \ell + 2$$

Since for any i with $2 \leq i \leq p-1$, we know that $p \mid \binom{p}{i}$, it follows that $p^{\ell+3}$ divides every term in the summation. Therefore,

$$(1 + ap^{\ell+1})^p \equiv 1 + ap^{\ell+2} \pmod{p^{\ell+3}}$$

Putting this together with the above, we conclude that

$$g^{p^{\ell+1}(p-1)} \equiv 1 + ap^{\ell+2} \pmod{p^{\ell+3}}$$

This completes the induction. □

Theorem 2.7.10. *Suppose that p is an odd prime. If g is a primitive root modulo p^2, then g is a primitive root modulo p^k for all $k \geq 2$.*

Proof. Let g be a primitive root modulo p^2 and let $k \geq 2$. Let n be the order of $\overline{g}$ in $U(\mathbb{Z}/p^k\mathbb{Z})$. Following the above arguments, we know that $n = p^m(p-1)$ for some $m \in \mathbb{N}$ with $0 \leq m \leq k-1$. We also know that we may write $g^{p-1} = 1 + ap$ for some $a \in \mathbb{Z}$ with $p \nmid a$. Now suppose that $m \leq k-2$, We then have $p^m(p-1) \mid p^{k-2}(p-1)$ and hence $g^{p^{k-2}(p-1)} \equiv 1 \pmod{p^k}$. However, the previous proposition tells us that

$$g^{p^{k-2}(p-1)} \equiv 1 + ap^{k-1} \pmod{p^k}$$

This would imply that $1 + ap^{k-1} \equiv 1 \pmod{p^k}$, contradicting the fact that $p \nmid a$. Therefore, we must have $m = k-1$, and thus $n = p^{k-1}(p-1)$. It follows that the order of $\overline{g}$ in $U(\mathbb{Z}/p^k\mathbb{Z})$ is $\varphi(p^k)$, so g is a primitive root modulo p^k. □

Corollary 2.7.11. *If p is an odd prime, then $U(\mathbb{Z}/p^k\mathbb{Z})$ is cyclic for all $k \in \mathbb{N}$.*

Suppose now that $n \geq 2$, and write the prime factorization of n as

$$n = p_1^{k_1} p_2^{k_2} \cdots p_\ell^{k_\ell}$$

By the Chinese Remainder Theorem, we know that

$$U(\mathbb{Z}/n\mathbb{Z}) \cong U(\mathbb{Z}/p_1^{k_1}\mathbb{Z}) \times U(\mathbb{Z}/p_1^{k_2}\mathbb{Z}) \times \cdots \times U(\mathbb{Z}/p_\ell^{k_\ell}\mathbb{Z})$$

Since we have classified all of the groups in the direct product on the right, we can use this to understand the structure of any $U(\mathbb{Z}/n\mathbb{Z})$.

Theorem 2.7.12. *Let $n \geq 2$. We then have that $U(\mathbb{Z}/n\mathbb{Z})$ is cyclic if and only if either $n = 2$, $n = 4$, $n = p^k$ for some odd prime p, or $n = 2p^k$ for some odd prime p.*

Proof. We know that $U(\mathbb{Z}/n\mathbb{Z})$ is cyclic in each of the first three cases. If $n = 2p^k$ for some odd prime p, then

$$U(\mathbb{Z}/n\mathbb{Z}) \cong U(\mathbb{Z}/2\mathbb{Z}) \times U(\mathbb{Z}/p^k\mathbb{Z}) \cong U(\mathbb{Z}/p^k\mathbb{Z})$$

because $U(\mathbb{Z}/2\mathbb{Z})$ is the trivial group, so $U(\mathbb{Z}/n\mathbb{Z})$ is cyclic.

Suppose conversely that n is not one of the above values. If $n = 2^k$ for some $k \geq 3$, then we've shown above that $U(\mathbb{Z}/n\mathbb{Z})$ is not cyclic. Otherwise, we can write $n = m_1 m_2$ where $\gcd(m_1, m_2) = 1$ and $m_1, m_2 \geq 3$. In this case, both $\varphi(m_1)$ and $\varphi(m_2)$ are even by the homework and

$$U(\mathbb{Z}/n\mathbb{Z}) \cong U(\mathbb{Z}/m_1\mathbb{Z}) \times U(\mathbb{Z}/m_2\mathbb{Z})$$

Since each of the groups on the right have even order, they each have elements of order 2. Thus, $U(\mathbb{Z}/n\mathbb{Z})$ has at least 2 elements of order 2, but a cyclic group has only $\varphi(2) = 1$ many elements of order 2. Therefore, $U(\mathbb{Z}/n\mathbb{Z})$ is not cyclic. □

2.8 When -1 is a Square Modulo p

Theorem 2.8.1. *Let p be an odd prime. There exists an $a \in \mathbb{Z}$ with $a^2 \equiv -1 \pmod{p}$ if and only if $p \equiv 1 \pmod{4}$.*

Algebraic Proof. Suppose first that there exists an $a \in \mathbb{Z}$ with $a^2 \equiv -1 \pmod{p}$. Working in $\mathbb{Z}/p\mathbb{Z}$, we then have that $\overline{a} \neq \overline{0}$ and $\overline{a} \neq \overline{1}$ (since $\overline{0}^2 = \overline{0} \neq \overline{-1}$ and $\overline{-1}^2 = \overline{1} \neq \overline{-1}$ as $p \geq 3$). In particular, $\overline{a} \in U(\mathbb{Z}/p\mathbb{Z})$ and $\overline{a}$ is not the identity of this group. We also have

$$\overline{a}^4 = (\overline{a^2})^2 = (\overline{-1})^2 = \overline{1}$$

Thus, the order of $\overline{a} \in \mathbb{Z}/p\mathbb{Z}$ is a divisor of 4. We know that this order is not 1 (since $\overline{a} \neq \overline{1}$) and it is also not 2 because $\overline{a}^2 = \overline{-1} \neq \overline{1}$. Therefore, the order of $\overline{a} \in U(\mathbb{Z}/p\mathbb{Z})$ is 4. By Lagrange's Theorem, it follows that $4 \mid \varphi(p)$, which is to say that $4 \mid p - 1$. Therefore, $p \equiv 1 \pmod{4}$.

Suppose conversely that $p \equiv 1 \pmod{4}$. We then have that $4 \mid p - 1$, so $4 \mid |U(\mathbb{Z}/p\mathbb{Z})|$. Recall that if G is a cyclic group of order n and d is a positive divisor of n, then G has an element of order d. Since $U(\mathbb{Z}/p\mathbb{Z})$ is a cyclic group and $4 \mid |U(\mathbb{Z}/p\mathbb{Z})|$, we may fix $a \in \mathbb{Z}$ with $p \nmid a$ such that $|\overline{a}| = 4$. We then have $\overline{a}^2 \neq \overline{1}$ and $(\overline{a}^2)^2 = \overline{a}^4 = \overline{1}$. Now the only solutions to $x^2 = 1$ in $\mathbb{Z}/p\mathbb{Z}$ are $\overline{1}$ and $\overline{-1}$, so we conclude that $\overline{a}^2 = \overline{-1}$. Therefore, $a^2 \equiv -1 \pmod{p}$. □

Elementary Proof. Suppose first that there exists an $a \in \mathbb{Z}$ with $a^2 \equiv -1 \pmod{p}$. Notice that $p \nmid a$ because otherwise $a^2 \equiv 0 \not\equiv -1 \pmod{p}$. By Fermat's Little Theorem, we have $a^{p-1} \equiv 1 \pmod{p}$. Thus

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

Since $p \geq 3$, we have $1 \not\equiv -1 \pmod{3}$ and thus $\frac{p-1}{2}$ must be even. It follows that $4 \mid (p-1)$, i.e. that $p \equiv 1 \pmod{4}$.

Suppose conversely that $p \equiv 1 \pmod{4}$. We then have that $4 \mid (p-1)$, so $\frac{p-1}{2}$ is even. Let $\ell = \frac{p-1}{2}$. By Wilson's Theorem we have

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

Now $-k \equiv p-k \pmod{p}$ for all k, so in the product $(p-1)!$ we can replace the latter half of the elements in the product $(p-1)!$ by the first ℓ negative numbers, i.e. the numbers in the list

$$\ell+1, \ell+2, \ldots, p-2, p-1$$

are each equivalent modulo p to exactly one number in the following list:

$$-\ell, -(\ell-1), \ldots, -2, -1$$

Therefore, working modulo p, we have

$$\begin{aligned}
-1 &\equiv (p-1)! \\
&\equiv 1 \cdot 2 \cdots (\ell-1) \cdot \ell \cdot (\ell+1) \cdot (\ell+2) \cdots (p-2) \cdot (p-1) \\
&\equiv 1 \cdot 2 \cdots (\ell-1) \cdot \ell \cdot (-\ell) \cdot (-(\ell-1)) \cdots (-2) \cdot (-1) \\
&\equiv (-1)^{\ell} \cdot 1 \cdot 2 \cdots (\ell-1) \cdot \ell \cdot \ell \cdot (\ell-1) \cdots 2 \cdot 1 \\
&\equiv (-1)^{\ell} \cdot (\ell!)^2 \\
&\equiv (\ell!)^2
\end{aligned}$$

where the last line follows because $\ell = \frac{p-1}{2}$ is even. Therefore, there exists $a \in \mathbb{Z}$ with $a^2 \equiv -1 \pmod{p}$, namely $a = \ell! = (\frac{p-1}{2})!$.

(A slightly cleaner way to write the above argument is simply to notice that you can list the elements of $U(\mathbb{Z}/p\mathbb{Z})$ as

$$\overline{-\ell}, \overline{-(\ell-1)} \ldots, \overline{-2}, \overline{-1}, \overline{1}, \overline{2}, \overline{\ell-1}, \overline{\ell}$$

rather the usual way as $\overline{1}, \overline{2}, \ldots, \overline{p-1}$, and then apply the above two propositions.) $\square$

Chapter 3

Abstracting the Integers

3.1 Euclidean Domains

Both $\mathbb{Z}$ and $F[x]$ (for F a field) have "division with remainder" properties. In both of these rings, the ability to divide and always get something "smaller" allows one to prove many powerful things. In particular, this property allowed us to prove the existence of greatest common divisors which eventually lead to unique factorization. With such power, we will define a class of integral domains based on the idea of allowing "division with remainder" so that our results will be as general as possible.

Definition 3.1.1. *Let R be an integral domain. A function $N\colon R\backslash\{0\} \to \mathbb{N}$ is called a Euclidean function on R if for all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that*

$$a = qb + r$$

and either $r = 0$ or $N(r) < N(b)$.

Definition 3.1.2. *An integral domain R is a* Euclidean domain *if there exists a Euclidean function on R.*

Example 3.1.3. *In algebra, you established the following.*

- *The function $N\colon \mathbb{Z}\backslash\{0\} \to \mathbb{N}$ defined by $N(a) = |a|$ is a Euclidean function on $\mathbb{Z}$, so $\mathbb{Z}$ is a Euclidean domain.*
- *Let F be a field. The function $N\colon F[x]\backslash\{0\} \to \mathbb{N}$ defined by $N(f(x)) = \deg(f(x))$ is a Euclidean function on $F[x]$, so $F[x]$ is a Euclidean domain.*

Notice that we do not require the uniqueness of q and r in our definition of a Euclidean function. Although it was certainly a nice perk to have some aspect of uniqueness in $\mathbb{Z}$ and $F[x]$, it turns out to be be unnecessary for the theoretical results of interest about Euclidean domains. Furthermore, many natural Euclidean functions on integral domains for which uniqueness fails, and we want to be as general as possible.

The name *Euclidean domain* comes from the fact that any such integral domain supports the ability to find greatest common divisors via the Euclidean algorithm. In particular, the notion of "size" given by a Euclidean function $N\colon R \to \mathbb{N}$ allows us to use induction to prove the existence of greatest common divisors. We begin with the following generalization of a simple result we proved about $\mathbb{Z}$ which works in any integral domain (even any commutative ring).

Proposition 3.1.4. *Let R be an integral domain. Let $a, b, q, r \in R$ with $a = qb + r$. For any $d \in R$, we have that d is a common divisor of a and b if and only if d is a common divisor of b and r, i.e.*

$$\{d \in R : d \text{ is a common divisor of } a \text{ and } b\} = \{d \in R : d \text{ is a common divisor of } b \text{ and } r\}$$

Proof. Suppose first that d is a common divisor of b and r. Since $d \mid b$, $d \mid r$, and $a = qb + r = bq + r1$, it follows that $d \mid a$.

Conversely, suppose that d is a common divisor of a and b. Since $d \mid a$, $d \mid b$, and $r = a - qb = a1 + b(-q)$, it follows that $d \mid r$. □

Theorem 3.1.5. *Let R be a Euclidean domain. Every pair of elements $a, b \in R$ has a greatest common divisor.*

Proof. Since R is a Euclidean domain, we may fix a Euclidean function $N\colon R\backslash\{0\} \to \mathbb{N}$. We use (strong) induction on $N(b) \in \mathbb{N}$ to prove the result. We begin by noting that if $b = 0$, then the set of common divisors of a and b equals the set of divisors of a (because every integer divides 0), so a satisfies the requirement of a greatest common divisor. Suppose then that $b \in R$ is nonzero and we know the result for all pairs $x, y \in R$ with either $y = 0$ or $N(y) < N(b)$. Fix $q, r \in R$ with $a = qb + r$ and either $r = 0$ or $N(r) < N(b)$. By (strong) induction, we know that b and r have a greatest common divisor d. By the Proposition 3.1.4, the set of common divisors of a and b equals the set of common divisors of b and r. It follows that d is a greatest common divisor of a and b. □

As an example, consider working in the ring $\mathbb{Q}[x]$ and trying to find a greatest common divisor of the following two polynomials:

$$f(x) = x^5 + 3x^3 + 2x^2 + 6 \qquad g(x) = x^4 - x^3 + 4x^2 - 3x + 3$$

We apply the Euclidean Algorithm as follows (we suppress the computations of the long divisions):

$$\begin{aligned} x^5 + 3x^3 + 2x^2 + 6 &= (x+1)(x^4 - x^3 + 4x^2 - 3x + 3) + (x^2 + 3) \\ x^4 - x^3 + 4x^2 - 3x + 3 &= (x^2 - x + 1)(x^2 + 3) + 0 \end{aligned}$$

Thus, the set of common of $f(x)$ and $g(x)$ equals the set of common divisors of $x^2 + 3$ and 0, which is just the set of divisors of $x^2 + 3$. Therefore, $x^2 + 3$ is a greatest common divisor of $f(x)$ and $g(x)$. Now this is not the only greatest common divisor because we know that any associate of $x^2 + 3$ will also be a greatest common divisor of $f(x)$ and $g(x)$. The units in $\mathbb{Q}[x]$ are the nonzero constants, so other greatest common divisors are $2x^2 + 6$, $\frac{5}{6}x^2 + \frac{5}{2}$, etc. We would like to have a canonical choice for which to pick, akin to choosing the nonnegative value when working in $\mathbb{Z}$.

Definition 3.1.6. *Let F be a field. A* monic *polynomial in $F[x]$ is a nonzero polynomial whose leading term is* 1.

Notice that every nonzero polynomial in $F[x]$ is an associate with a unique monic polynomial (if the leading term is $a \neq 0$, just multiply by a^{-1} to get a monic associate, and notice that this is the only way to multiply by a nonzero constant to make it monic). By restricting to monic polynomials, we get a canonical choice for a greatest common divisor.

Definition 3.1.7. *Let F be a field and let $f(x), g(x) \in F[x]$ be polynomials. If at least one of $f(x)$ and $g(x)$ is nonzero, we define* $\gcd(f(x), g(x))$ *to be the unique monic polynomial which is a greatest common divisor of $f(x)$ and $g(x)$. Notice that if both $f(x)$ and $g(x)$ are the zero polynomial, then 0 is the only greatest common divisor of $f(x)$ and $g(x)$, so we define* $\gcd(f(x), g(x)) = 0$.

Now $x^2 + 3$ is monic, so from the above computations, we have

$$\gcd(x^5 + 3x^3 + 2x^2 + 6, x^4 - x^3 + 4x^2 - 3x + 3) = x^2 + 3$$

We now provide another incredibly important an example by showing that the Gaussian Integers $\mathbb{Z}[i]$ are also a Euclidean domain.

Definition 3.1.8. *Working in the field $\mathbb{C}$, we define the following.*

- $\mathbb{Q}(i) = \{q + ri : q, r \in \mathbb{Q}\}$
- $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$

Notice that $\mathbb{Q}(i)$ is a subfield of $\mathbb{C}$ and $\mathbb{Z}[i]$ is a subring of $\mathbb{Q}(i)$ and thus of $\mathbb{C}$. The ring $\mathbb{Z}[i]$ is called the Gaussian Integers.

To see that $\mathbb{Q}(i)$ is a field, suppose that $\alpha \in \mathbb{Q}(i)$ is nonzero and write $\alpha = q + ri$. We then have that either $q \neq 0$ or $r \neq 0$, so

$$\begin{aligned}\frac{1}{\alpha} &= \frac{1}{q+ri} \\ &= \frac{1}{q+ri} \cdot \frac{q-ri}{q-ri} \\ &= \frac{q-ri}{q^2+r^2} \\ &= \frac{q}{q^2+r^2} + \frac{-r}{q^2+r^2} \cdot i\end{aligned}$$

Since both $\frac{q}{q^2+r^2}$ and $\frac{-r}{q^2+r^2}$ are elements of $\mathbb{Q}$, it follows that $\frac{1}{\alpha} \in \mathbb{Q}(i)$.

Definition 3.1.9. *We define a function $N\colon \mathbb{Q}(i) \to \mathbb{Q}$ by letting $N(q + ri) = q^2 + r^2$. The function N is called the* norm *on the field $\mathbb{Q}(i)$.*

Proposition 3.1.10. *For the function $N(q + ri) = q^2 + r^2$ defined on $\mathbb{Q}(i)$, we have*

1. *$N(\alpha) \geq 0$ for all $\alpha \in \mathbb{Q}(i)$.*
2. *$N(\alpha) = 0$ if and only if $\alpha = 0$.*
3. *$N(q) = q^2$ for all $q \in \mathbb{Q}$.*
4. *$N(\alpha) \in \mathbb{N}$ for all $\alpha \in \mathbb{Z}[i]$.*
5. *$N(\alpha\beta) = N(\alpha) \cdot N(\beta)$ for all $\alpha, \beta \in \mathbb{Q}(i)$.*

Proof. The first four are all immediate from the definition. Suppose that $\alpha, \beta \in \mathbb{Q}(i)$ and write $\alpha = q + ri$ and $\beta = s + ti$. We have

$$\begin{aligned}N(\alpha\beta) &= N((q+ri)(s+ti)) \\ &= N(qs + rsi + qti - rt) \\ &= N((qs - rt) + (rs + qt)i) \\ &= (qs - rt)^2 + (rs + qt)^2 \\ &= q^2s^2 - 2qsrt + r^2t^2 + r^2s^2 + 2rsqt + q^2t^2 \\ &= q^2s^2 + r^2s^2 + q^2t^2 + r^2t^2 \\ &= (q^2 + r^2) \cdot (s^2 + t^2) \\ &= N(q + ri) \cdot N(s + ti) \\ &= N(\alpha) \cdot N(\beta)\end{aligned}$$

□

We first use this result to classify the units of $\mathbb{Z}[i]$.

Proposition 3.1.11. *We have $U(\mathbb{Z}[i]) = \{1, -1, i, -i\}$.*

Proof. We have $1 \cdot 1 = 1$, $(-1) \cdot (-1) = 1$, and $i \cdot (-i) = 1$, so all of these elements are units. Suppose now that $\alpha \in \mathbb{Z}[i]$ is a unit and fix $\beta \in \mathbb{Z}[i]$ with $\alpha\beta = 1$. We then have

$$\begin{aligned} 1 &= 1^2 + 0^2 \\ &= N(1) \\ &= N(\alpha\beta) \\ &= N(\alpha) \cdot N(\beta) \end{aligned}$$

Since $N(\alpha)$ and $N(\beta)$ are integers, it follows that $N(\alpha) \mid 1$ in $\mathbb{Z}$. Using the fact that $N(\alpha) \geq 0$, it follows that $N(\alpha) = 1$. Write $\alpha = a + bi$ where $a, b \in \mathbb{Z}$. We then have

$$1 = N(\alpha) = a^2 + b^2$$

We conclude that $|a| \leq 1$ and $|b| \leq 1$. Since $a, b \in \mathbb{Z}$, we can work through the various possibilities to deduce that (a, b) is one of the following pairs: $(1, 0)$, $(-1, 0)$, $(0, 1)$, or $(0, -1)$. Therefore, $\alpha \in \{1, -1, i, -i\}$. □

Theorem 3.1.12. *$\mathbb{Z}[i]$ is a Euclidean domain with Euclidean function $N(a + bi) = a^2 + b^2$.*

Proof. We already know that $\mathbb{Z}[i]$ is an integral domain. Suppose that $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$. When we divide α by β in the field $\mathbb{Q}(i)$ we get $\frac{\alpha}{\beta} = s + ti$ for some $s, t \in \mathbb{Q}$. Fix integers $m, n \in \mathbb{Z}$ closest to $s, t \in \mathbb{Q}$ respectively, i.e. fix $m, n \in \mathbb{Z}$ so that $|m - s| \leq \frac{1}{2}$ and $|n - t| \leq \frac{1}{2}$. Let $\gamma = m + ni \in \mathbb{Z}[i]$, and let $\rho = \alpha - \beta\gamma \in \mathbb{Z}[i]$. We then have that $\alpha = \beta\gamma + \rho$, so we need only show that $N(\rho) < N(\beta)$. Now

$$\begin{aligned} N(\rho) &= N(\alpha - \beta\gamma) \\ &= N(\beta \cdot (s + ti) - \beta \cdot \gamma) \\ &= N(\beta \cdot ((s + ti) - (m + ni))) \\ &= N(\beta \cdot ((s - m) + (t - n)i)) \\ &= N(\beta) \cdot N((s - m) + (t - n)i) \\ &= N(\beta) \cdot ((s - m)^2 + (t - n)^2) \\ &\leq N(\beta) \cdot (\frac{1}{4} + \frac{1}{4}) \\ &= \frac{1}{2} \cdot N(\beta) \\ &< N(\beta) \end{aligned}$$

where the last line follows because $N(\beta) > 0$. □

We work out an example of finding a greatest common of $8 + 9i$ and $10 - 5i$ in $\mathbb{Z}[i]$. We follow the proof to find quotients and remainders. Notice that

$$\begin{aligned} \frac{8 + 9i}{10 - 5i} &= \frac{8 + 9i}{10 - 5i} \cdot \frac{10 + 5i}{10 + 5i} \\ &= \frac{80 + 40i + 90i - 45}{100 + 25} \\ &= \frac{35 + 130i}{125} \\ &= \frac{7}{25} + \frac{26}{25} \cdot i \end{aligned}$$

Following the proof (where we take the closest integers to $\frac{7}{25}$ and $\frac{26}{25}$), we should use the quotient i and determine the remainder from there. We thus write

$$8 + 9i = i \cdot (10 - 5i) + (3 - i)$$

Notice that $N(3 - i) = 9 + 1 = 10$ which is less than $N(10 - 5i) = 100 + 25 = 125$. Following the Euclidean algorithm, we next calculate

$$\begin{aligned}\frac{10 - 5i}{3 - i} &= \frac{10 - 5i}{3 - i} \cdot \frac{3 + i}{3 + i} \\ &= \frac{30 + 10i - 15i + 5}{9 + 1} \\ &= \frac{35 - 5i}{10} \\ &= \frac{7}{2} - \frac{1}{2} \cdot i\end{aligned}$$

Following the proof (where we now have many choices because $\frac{7}{2}$ is equally close to 3 and 4 and $-\frac{1}{2}$ is equally close to -1 and 0), we choose to take the quotient 3. We then write

$$10 - 5i = 3 \cdot (3 - i) + (1 - 2i)$$

Notice that $N(1 - 2i) = 1 + 4 = 5$ which is less than $N(3 - i) = 9 + 1 = 10$. Going to the next step, we calculate

$$\begin{aligned}\frac{3 - i}{1 - 2i} &= \frac{3 - i}{1 - 2i} \cdot \frac{1 + 2i}{1 + 2i} \\ &= \frac{3 + 6i - i + 2}{1 + 4} \\ &= \frac{5 + 5i}{5} \\ &= 1 + i\end{aligned}$$

Therefore, we have

$$3 - i = (1 + i) \cdot (1 - 2i) + 0$$

Putting together the various divisions, we see the Euclidean algorithm as:

$$\begin{aligned}8 + 9i &= i \cdot (10 - 5i) + (3 - i) \\ 10 - 5i &= 3 \cdot (3 - i) + (1 - 2i) \\ 3 - i &= (1 + i) \cdot (1 - 2i) + 0\end{aligned}$$

Thus, the set of common divisors of $8 + 9i$ and $10 - 5i$ equals the set of common divisors of $1 - 2i$ and 0, which is just the set of divisors of $1 - 2i$. Since a greatest common divisor is unique up to associates and the units of $\mathbb{Z}[i]$ are $1, -1, i, -i$, it follows the set of greatest common divisors of $8 + 9i$ and $10 - 5i$ is

$$\{1 - 2i, -1 + 2i, 2 + i, -2 - i\}$$

3.2 Principal Ideal Domains

We chose our definition of a Euclidean domain to abstract away the fundamental fact about $\mathbb{Z}$ that we can always divide in such a way to get a quotient along with a "smaller" remainder. As we have seen, this ability

allows us to carry over to these more general rings the existence of greatest common divisors and the method of finding them via the Euclidean Algorithm.

Recall back when we working with $\mathbb{Z}$ that we had another characterization of (and proof of existence for) the greatest common divisor. We proved that the greatest common divisor of two nonzero integers a and b was the least positive number of the form $ma + nb$ where $m, n \in \mathbb{Z}$. Now the "least" part will have no analogue in a general integral domain, so we will have to change that. Perhaps surprisingly, it turns out that the way to generalize this construction is to work with ideals. As we will see, in hindsight, what makes this approach to greatest common divisors work in $\mathbb{Z}$ is the fact that every ideal of $\mathbb{Z}$ is principal. We give the integral domains which have this property a special name.

Definition 3.2.1. *A* principal ideal domain*, or* PID*, is an integral domain in which every ideal is principal.*

Before working with these rings on their own terms, we first prove that every Euclidean domains is a PID so that we have a decent supply of examples. Our proof generalizes the one for $\mathbb{Z}$ in the sense that instead of looking for a smallest positive element of the ideal we simply look for an element of smallest "size" according to a given Euclidean function.

Theorem 3.2.2. *Every Euclidean domain is a PID.*

Proof. Let R be a Euclidean domain, and fix a Euclidean function $N\colon R\backslash\{0\} \to \mathbb{N}$. Suppose that I is an ideal of R. If $I = \{0\}$, then $I = \langle 0\rangle$. Suppose then that $I \neq \{0\}$. The set

$$\{N(a) : a \in I\backslash\{0\}\}$$

is a nonempty subset of $\mathbb{N}$. By the well-ordering property of $\mathbb{N}$, the set has a least element m. Fix $b \in I$ with $N(b) = m$. Since $b \in I$, we clearly have $\langle b\rangle \subseteq I$. Suppose now that $a \in I$. Fix $q, r \in R$ with

$$a = qb + r$$

and either $r = 0$ or $N(r) < N(b)$. Since $r = a - qb$ and both $a, d \in I$, it follows that $r \in I$. Now if $r \neq 0$, then $N(r) < N(b) = m$ contradicting our minimality of m. Therefore, we must have $r = 0$ and so $a = qb$. It follows that $a \in \langle b\rangle$. Since $a \in I$ was arbitrary, we conclude that $I \subseteq \langle b\rangle$. Therefore, $I = \langle b\rangle$. □

Corollary 3.2.3. *$\mathbb{Z}$, $F[x]$ for F a field, and $\mathbb{Z}[i]$ are all PIDs.*

Notice also that all fields F are also PIDs for the trivial reason that the only ideals of F are $\{0\} = \langle 0\rangle$ and $F = \langle 1\rangle$. In fact, all fields are also trivially Euclidean domain via absolutely any function $N\colon F\backslash\{0\} \to \mathbb{N}$ because you can always divide by a nonzero element with zero as a remainder.

It turns out that there are PIDs which are not Euclidean domains, but we will not construct examples of such rings now. Returning to our other characterization of greatest common divisors in $\mathbb{Z}$, we had that if $a, b \in \mathbb{Z}$ not both nonzero, then we considered the set

$$\{ma + nb : m, n \in \mathbb{Z}\}$$

and proved that the least positive element of this set was the greatest common divisor. In our current ring-theoretic language, the above set is the ideal $\langle a, b\rangle$ of $\mathbb{Z}$, and a generator of this ideal is a greatest common divisor. With this change in perspective/language, we can carry this argument over to an arbitrary PID.

Theorem 3.2.4. *Let R be a PID and let $a, b \in R$.*

1. *There exists a greatest common divisor of a and b.*
2. *If d is a greatest common divisor of a and b, then there exists $r, s \in R$ with $d = ra + sb$.*

Proof.

1. Let $a, b \in R$. Consider the ideal

$$I = \langle a, b \rangle = \{ra + sb : r, s \in R\}$$

Since R is a PID, the ideal I is principal, so we may fix $d \in R$ with $I = \langle d \rangle$. Since $d \in \langle d \rangle = \langle a, b \rangle$, we may fix $r, s \in R$ with $ra + sb = d$. We claim that d is a greatest common divisor of a and b.

First notice that $a \in I$ since $a = 1a + 0b$, so $a \in \langle d \rangle$, and hence $d \mid a$. Also, we have $b \in I$ because $b = 0a + 1b$, so $b \in \langle d \rangle$, and hence $d \mid b$. Thus, d is a common divisor of a and b.

Suppose now that c is a common divisor of a and b. Fix $m, n \in R$ with $a = cm$ and $b = cn$. We then have

$$\begin{aligned} d &= ra + sb \\ &= r(cm) + s(cn) \\ &= c(rm + sn) \end{aligned}$$

Thus, $c \mid d$. Putting it all together, we conclude that d is a greatest common divisor of a and b.

2. For the d in part 1, we showed in the proof that there exist $r, s \in R$ with $d = ra + sb$. Let d' be any other greatest common divisor of a and b, and fix a unit u with $d' = du$. We then have

$$d' = du = (ra + sb)u = a(ru) + b(su)$$

□

If you are given $a, b \in R$ and you know a greatest common divisor d of a and b, how can you explicitly calculate $r, s \in R$ with $ra + sb = d$? In a general PID, this can be very hard. However, suppose you are in the special case where R is a Euclidean domain. Assuming that we can explicitly calculate quotients and remainders for repeated division (as we could in $\mathbb{Z}$, $F[x]$, and $\mathbb{Z}[i]$), we can calculate a greatest common divisor d of a and b by "winding up" the Euclidean algorithm backwards as in $\mathbb{Z}$.

For example, working in the Euclidean domain $\mathbb{Z}[i]$, we computed in the last section that $1 - 2i$ is a greatest common divisor of $8 + 9i$ and $10 - 5i$ by applying the Euclidean algorithm to obtain:

$$\begin{aligned} 8 + 9i &= i \cdot (10 - 5i) + (3 - i) \\ 10 - 5i &= 3 \cdot (3 - i) + (1 - 2i) \\ 3 - i &= (1 + i) \cdot (1 - 2i) + 0 \end{aligned}$$

Working backwards, we see that

$$\begin{aligned} 1 - 2i &= 1 \cdot (10 - 5i) + (-3) \cdot (3 - i) \\ &= 1 \cdot (10 - 5i) + (-3) \cdot [(8 + 9i) - i \cdot (10 - 5i)] \\ &= (1 + 3i) \cdot (10 - 5i) + (-3) \cdot (8 + 9i) \end{aligned}$$

It is possible to define a greatest common divisor of elements $a_1, a_2, \ldots, a_n \in R$ completely analogously to our definition for pairs of elements. If you do so, even in the case of a nice Euclidean domain, you can not immediately generalize the idea of the Euclidean Algorithm to many elements without doing a kind of repeated nesting that gets complicated. However, notice that you can very easily generalize our PID arguments to prove that greatest common divisors exist and are unique up to associates by following the above proofs and simply replacing the ideal $\langle a, b \rangle$ with the ideal $\langle a_1, a_2, \ldots, a_n \rangle$. You even conclude that it possible to write a greatest common divisor in the form $r_1 a_1 + r_2 a_2 + \cdots + r_n a_n$. The assumption that all ideals are principal is extremely powerful.

With the hard work of the last couple of sections in hand, we can now carry over much of our later work in $\mathbb{Z}$ which dealt with relatively prime integers and primes. The next definition and ensuing two propositions directly generalize corresponding results about $\mathbb{Z}$.

Definition 3.2.5. *Let R be a PID. Two elements $a, b \in R$ are* relatively prime *if 1 is a greatest common divisor of a and b.*

Proposition 3.2.6. *Let R be a PID and let $a, b, c \in R$. If $a \mid bc$ and a and b are relatively prime, then $a \mid c$.*

Proof. Fix $d \in R$ with $bc = ad$. Fix $r, s \in R$ with $ra + sb = 1$. Multiplying this last equation through by c, we conclude that $rac + sbc = c$, so

$$\begin{aligned} c &= rac + s(bc) \\ &= rac + s(ad) \\ &= a(rc + sd) \end{aligned}$$

It follows that $a \mid c$. □

Proposition 3.2.7. *Suppose that R is a PID. If p is irreducible, then p is prime.*

Proof. Suppose that $p \in R$ is irreducible. By definition, p is nonzero and not a unit. Suppose that $a, b \in R$ are such that $p \mid ab$. Fix a greatest common divisor d of p and a. Since $d \mid p$, we may fix $c \in R$ with $p = dc$. Now p is irreducible, so either d is a unit or c is a unit. We handle each case.

- Suppose that d is a unit. We then have that 1 is an associate of d, so 1 is also a greatest common divisor of p and a. Therefore, p and a are relatively prime, so as $p \mid ab$ we may use the previous corollary to conclude that $p \mid b$.
- Suppose that c is a unit. We then have that $pc^{-1} = d$, so $p \mid d$. Since $d \mid a$, it follows that $p \mid a$.

Therefore, either $p \mid a$ or $p \mid b$. It follows that p is prime. □

Definition 3.2.8. *Let R be a commutative ring.*

- *A* prime ideal *of a ring R is an ideal $P \neq R$ such that whenever $ab \in P$, either $a \in P$ or $b \in P$.*
- *A* maximal ideal *of a ring R is an ideal $M \neq R$ such that there is no ideal I with $M \subsetneq I \subsetneq R$.*

Proposition 3.2.9. *Let R be a commutative ring and let $p \in R$ be nonzero. The ideal $\langle p \rangle$ is a prime ideal of R if and only if p is a prime element of R.*

Proof. Suppose first that $\langle p \rangle$ is a prime ideal of R. Notice that $p \neq 0$ by assumption and that p is not a unit because $\langle p \rangle \neq R$. Suppose that $a, b \in R$ and $p \mid ab$. We then have that $ab \in \langle p \rangle$, so as $\langle p \rangle$ is a prime ideal we know that either $a \in \langle p \rangle$ or $b \in \langle p \rangle$. In the former case, we conclude that $p \mid a$, and in the latter case we conclude that $p \mid b$. Since $a, b \in R$ were arbitrary, it follows that p is a prime element of R.

Suppose conversely that p is a prime element of R. Suppose that $a, b \in R$ and $ab \in \langle p \rangle$. We then have that $p \mid ab$, so as p is a prime element we know that either $p \mid a$ or $p \mid b$. In the former case, we conclude that $a \in \langle p \rangle$ and in the latter case we conclude that $b \in \langle p \rangle$. Since $a, b \in R$ were arbitrary, it follows that $\langle p \rangle$ is a prime ideal of R. □

Recall the following important theorem from algebra.

Theorem 3.2.10. *Let I be an ideal of the commutative ring R.*

- *I is a prime ideal of R if and only if R/I is an integral domain.*
- *I is a maximal ideal of R if and only if R/I is a field.*

Corollary 3.2.11. *In a commutative ring R, every maximal ideal is a prime ideal.*

Proposition 3.2.12. *Let R be a PID and let $a \in R$ be nonzero. The following are equivalent.*

1. *$\langle a \rangle$ is a maximal ideal.*
2. *$\langle a \rangle$ is a prime ideal.*
3. *a is a prime.*
4. *a is irreducible.*

Proof. Notice that $1 \to 2 \to 3 \to 4$ from the above results (or see Section 10.5 of the Algebra notes). We prove that $4 \to 1$. Suppose that a is irreducible, and let $M = \langle a \rangle$. Since a is not a unit, we have that $1 \notin \langle a \rangle$, so $M \neq R$. Suppose that I is an ideal with $M \subseteq I \subseteq R$. Since R is a PID, there exists $b \in R$ with $I = \langle b \rangle$. We then have that $\langle a \rangle \subseteq \langle b \rangle$, so $b \mid a$. Fix $c \in R$ with $a = bc$. Since a is irreducible, either b is a unit or c is a unit. In the former case, we have $I = \langle b \rangle = R$, and in the latter case we have that b is an associate of a so $I = \langle b \rangle = \langle a \rangle = M$. □

Corollary 3.2.13. *If R is a PID, then every nonzero prime ideal is maximal.*

Notice that we need to assume that the ideal is nonzero because $\{0\}$ is a prime ideal of $\mathbb{Z}$ that is not maximal. In fact, in every integral domain that is that a field, $\{0\}$ is nonmaximal prime ideal.

3.3 Factorizations and Noetherian Rings

We now seek to seek to prove an analogue of the Fundamental Theorem of Arithmetic in "nice" integral domains (say Euclidean domains or even PIDs). We have developed most of the tools to do this when we showed that irreducibles are prime in any PID. This fact in $\mathbb{Z}$ (that if p is "prime" in $\mathbb{Z}$ and $p \mid ab$, then either $p \mid a$ or $p \mid d$) was the key tool used to establish the Fundamental Theorem of Arithmetic. However, there is one other aspect of the Fundamental Theorem of Arithmetic that is easy to overlook. Namely, why does even exist a factorization into irreducibles of every (nonzero nonunit) element? In $\mathbb{Z}$, this is easy to argue by induction because proper factors are smaller in absolute value. In Euclidean domains with "nice" Euclidean functions (say where $N(b) < N(a)$ whenever $b \mid a$ and b is not an associate of a), you can mimic this argument. For example, it is relatively straightforward to prove that in $\mathbb{Z}[i]$ every (nonzero nonunit) element factors into irreducibles. But what happens in a general PID?

Let's analyze where a problem could arise. Suppose R is an integral domain and that $a \in R$ is nonzero and not a unit. If a is irreducible, we have trivially factored it into irreducibles. Suppose then that a is not irreducible and write $a = bc$ where neither b nor c is a unit. If either b or c is irreducible, we are happy with that piece, but we may need to factor b or c (or perhaps both) further. And from here, their factors may break up further still. How do we know that this process of repeatedly breaking down element must eventually stop? There does not seem to be any strong reason to believe that it does if we do not have a way of saying that factors are "smaller". In fact, there are integral domains where some elements are not products of irreducibles and so this process goes on forever. However, this process must stop in any PID, and we go about developing the tools to prove that now. The key idea is to turn everything into ideals because that is the only aspect of a PID we have control over. Recall the following basic facts from algebra.

Proposition 3.3.1. *Let R be a commutative ring. For any $a, b \in R$ we have $a \mid b$ if and only if $\langle b \rangle \subseteq \langle a \rangle$.*

Proof. First notice that if $a \mid b$, then $b \in \langle a \rangle$, hence $\langle b \rangle \subseteq \langle a \rangle$ because $\langle b \rangle$ is the smallest ideal containing b. If conversely we have $\langle b \rangle \subseteq \langle a \rangle$, then clearly $b \in \langle a \rangle$, so $a \mid b$. □

Corollary 3.3.2. *Let R be an integral domain. For any $a, b \in R$, we have $\langle a \rangle = \langle b \rangle$ if and only if a and b are associates.*

Proof. If $\langle a\rangle = \langle b\rangle$, then both $a \mid b$ and $b \mid a$ by the previous proposition, so a and b are associates. Conversely, if a and b are associates, then both $a \mid b$ and $b \mid a$, hence $\langle a\rangle = \langle b\rangle$ from the previous proposition. □

Corollary 3.3.3. *Suppose that R is an integral domain and $a, b \in R$. If $b \mid a$ and b is not an associate of a, then $\langle a\rangle \subsetneq \langle b\rangle$.*

Proof. This follows immediately from the previous two results. □

Suppose now that R is an integral and $a \in R$ is nonzero and not a unit. Write $a = bc$ where b and c are both not units. We then have that neither b nor c is an associate of a, so $\langle a\rangle \subsetneq \langle b\rangle$ and $\langle a\rangle \subsetneq \langle c\rangle$. Thus, in terms of ideals, we have become larger. If we factor further still, we get even larger ideals. So we have turned the question on its head and need to think about whether this process of ever larger ideals can go on forever. We define a special class of rings where this does not happen.

Definition 3.3.4. *A commutative ring R is said to be Noetherian if whenever*

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$$

is a sequence of ideals, there exists $N \in \mathbb{N}$ such that $I_k = I_N$ for all $k \geq N$. Equivalently, there is no strictly increasing sequence of ideals

$$I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \dots$$

Proposition 3.3.5. *Let R be a commutative ring. We then have that R is Noetherian if and only if every ideal of R is finitely generated (i.e. for every ideal I of R, there exist $a_1, a_2, \dots, a_m \in R$ with $I = \langle a_1, a_2, \dots, a_m\rangle$).*

Proof. Suppose first that every ideal of R is finitely generated. Let

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$$

be a sequence of ideals. Let

$$J = \bigcup_{k=1}^{\infty} I_k = \{r \in R : r \in I_k \text{ for some } k \in \mathbb{N}^+\}$$

We claim that J is an ideal of R. First notice that $0 \in I_1 \subseteq J$. Suppose that $a, b \in J$. Fix $k, \ell \in \mathbb{N}^+$ with $a \in I_k$ and $b \in I_\ell$. We then have $a, b \in I_{\max\{k,\ell\}} \subseteq J$. Suppose that $a \in J$ and $r \in R$. Fix $k \in \mathbb{N}^+$ with $a \in I_k$. We then have that $ra \in I_k \subseteq J$.

Since J is an ideal of R and we are assuming that every ideal of R is finitely generated, there exist $a_1, a_2, \dots, a_m \in R$ with $J = \langle a_1, a_2, \dots, a_m\rangle$. For each i, fix $k_i \in \mathbb{N}$ with $a_i \in I_{k_i}$. Let $N = \max\{k_1, k_2, \dots, k_m\}$. We then have that $a_i \in I_N$ for each i, hence $J = \langle a_1, a_2, \dots, a_m\rangle \subseteq I_N$. Therefore, for any $n \geq N$, we have

$$I_N \subseteq I_n \subseteq J \subseteq I_N$$

hence $I_n = I_N$.

Suppose conversely that some ideal of R is not finitely generated and fix such an ideal J. Define a sequence of elements of J as follows. Let a_1 be an arbitrary element of J. Suppose that we have defined $a_1, a_2, \dots, a_k \in J$. Since J is not finitely generated, we have that

$$\langle a_1, a_2, \dots, a_n\rangle \subsetneq J$$

so we may let a_{k+1} be some (any) element of $J \backslash \langle a_1, a_2, \dots, a_k\rangle$. Letting $I_n = \langle a_1, a_2, \dots, a_n\rangle$, we then have

$$I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \dots$$

so R is not Noetherian. □

Corollary 3.3.6. *Every PID is Noetherian.*

Proof. This follows immediately from the fact that in a PID every ideal is generated by one element. □

Theorem 3.3.7. *In a Noetherian integral domain, every nonzero nonunit element can be written as a product of irreducibles.*

Definition 3.3.8. *Let $\{0,1\}^*$ be the set of all finite sequences of 0's and 1's (including the "empty string" λ). A* tree *is a subset $T \subseteq \{0,1\}^*$ which is closed under initial segments. In other words, if $\sigma \in T$ and τ is an initial segment of S, then $\tau \in S$.*

For example, the set $\{\lambda, 0, 1, 00, 01, 011, 0110, 0111\}$ is a tree.

Lemma 3.3.9 (König's Lemma)**.** *Every infinite tree has an infinite branch. In other words, if T is a tree with infinitely many elements, then there is an infinite sequence of 0's and 1's such that every finite initial segment of this sequence is an element of T.*

Proof. Let T be a tree with infinitely many elements. We build the infinite sequences in stages. That is, we define finite sequences $\sigma_0 \prec \sigma_1 \prec \sigma_2 \prec \dots$ recursively where each $|\sigma_n| = n$. In our construction, we maintain the invariant that there are infinitely many element of T extending σ_n.

We begin by defining $\sigma_0 = \lambda$ and notice that there are infinitely many element of T extending λ because λ is an initial segment of every element of T trivially. Suppose that we have defined σ_n in such a way that $|\sigma_n| = n$ and there are infinitely many elements of T extending σ_n. We then must have that either there are infinitely many elements of T extending $\sigma_n 0$ or there are infinitely many elements of T extending $\sigma_n 1$. Thus, we may fix an $i \in \{0,1\}$ such that there are infinitely many elements of T extending $\sigma_n i$, and define $\sigma_{n+1} = \sigma_n i$.

We now take the unique infinite sequence extending all of the σ_n and notice that it has the required properties. □

Proof 1 of Theorem 3.3.7 - With König's Lemma. Suppose that $a \in R$ is a nonzero nonunit. Recursively factor a into nonunits down a tree stopping the growth at any irreducibles you reach. This tree can not have an infinite path because this would give a strictly increasing sequence of principal ideals. Therefore, by K"onig's Lemma, the tree is finite. It follows that a is the product of the leaves, and hence a product of irreducibles. □

Proof 2 of Theorem 3.3.7 - Without König's Lemma. Suppose that $a \in R$ is a nonzero nonunit which is not a product of irreducible elements. We define a sequence of elements $d_1, d_2, \dots$ in R as follows. Start by letting $d_1 = a$. Assume inductively that d_n is a nonzero nonunit which is not a product of irreducibles. In particular, d_n is itself not irreducible, so we may write $d_n = bc$ for some choice of nonzero nonunits b and c. Now it is not possible that both b and c are products of irreducibles because otherwise d_n would be as well. Thus, we may let d_{n+1} be one of b and c, chosen so that d_{n+1} is also not a product of irreducibles. Notice that d_{n+1} is a nonzero nonunit, that $d_{n+1} \mid d_n$, and d_{n+1} is not an associate of d_n because neither b nor c are units. Therefore,

$$\langle d_1 \rangle \subsetneq \langle d_2 \rangle \subsetneq \dots$$

contradicting the above proposition. It follows that every nonzero nonunit is a product of irreducibles. □

Corollary 3.3.10. *In every PID, every nonzero nonunit element can be written as a product of irreducibles.*

Proof. Immediate from the fact that every PID is Noetherian. □

We existence out of the way, we now move on to uniqueness of factorization into irreducibles. We try to generalize the results about the integers as much as possible.

Definition 3.3.11. *Let R be an integral domain and $p \in R$ be irreducible. Define a function $ord_p \colon R \to \mathbb{N} \cup \{\infty\}$ as follows. Given $a \in R$, let $ord_p(a)$ be the largest $k \in \mathbb{N}$ such that $p^k \mid a$ if one exists, and otherwise let $ord_p(a) = \infty$.*

Lemma 3.3.12. *Let R be a PID. Let $p \in R$ be irreducible, let $a \in R$, and let $k \in \mathbb{N}$. The following are equivalent.*

1. *$ord_p(a) = k$*
2. *$p^k \mid a$ and $p^{k+1} \nmid a$*
3. *There exists $m \in R$ with $a = p^k m$ and $p \nmid m$*

Proof.

- $1 \to 2$ is immediate.
- $2 \to 1$: Suppose that $p^k \mid a$ and $p^{k+1} \nmid a$. We clearly have $ord_p(a) \geq k$. Suppose that there exists $\ell > k$ with $p^\ell \mid a$. Since $\ell > k$, we have $\ell \geq k+1$. This implies that $p^{k+1} \mid p^\ell$, so since $p^\ell \mid a$ we conclude that $p^{k+1} \mid a$. This contradicts our assumption. Therefore, there is no $\ell > k$ with $p^\ell \mid a$, and hence $ord_p(a) = k$.
- $2 \to 3$: Suppose that $p^k \mid a$ and $p^{k+1} \nmid a$. Fix $m \in R$ with $a = p^k m$. If $p \mid m$, then we may fix $n \in R$ with $m = pn$, which would imply that $a = p^k pn = p^{k+1} n$ contradicting the fact that $p^{k+1} \nmid a$. Therefore, we must have $p \nmid m$.
- $3 \to 2$: Fix $m \in R$ with $a = p^k m$ and $p \nmid m$. We clearly have $p^k \mid a$. Suppose that $p^{k+1} \mid a$ and fix $n \in R$ with $a = p^{k+1} n$. We then have $p^k m = p^{k+1} n$, so $m = pn$. This implies that $p \mid m$, which is a contradiction. Therefore, $p^{k+1} \nmid a$.

□

The following theorem was essential in proving the Fundamental Theorem of Arithmetic. As usual, the key fact is that irreducibles are prime in PIDs.

Theorem 3.3.13. *Let R be a PID. Let $p \in R$ be irreducible. We have the following.*

1. *$ord_p(ab) = ord_p(a) + ord_p(b)$ for all $a, b \in R$.*
2. *$ord_p(a^n) = n \cdot ord_p(a)$ for all $a \in R$ and $n \in \mathbb{N}^+$.*

Proof. We follow the proofs in Homework 1. Let $a, b \in R$. First notice that if $ord_p(a) = \infty$, then $p^k \mid a$ for all $k \in \mathbb{N}$, hence $p^k \mid ab$ for all $k \in \mathbb{N}$, and thus $ord_p(ab) = \infty$. Similarly, if $ord_p(b) = \infty$, then $ord_p(ab) = \infty$. Suppose then that both $ord_p(a)$ and $ord_p(b)$ are finite, and let $k = ord_p(a)$ and $\ell = ord_p(b)$. Using Lemma 3.3.12, we may then write $a = p^k m$ where $p \nmid m$ and $b = p^\ell n$ where $p \nmid n$. We then have

$$ab = p^k m p^\ell n = p^{k+\ell} \cdot mn$$

Now if $p \mid mn$, then since p is prime (as it is irreducible and we are in a PID), we conclude that either $p \mid m$ or $p \mid n$, but both of these are contradictions. Therefore, $p \nmid mn$. Using Lemma 1.6 again, it follows that $ord_p(ab) = k + \ell$. □

Lemma 3.3.14. *Let R be a PID, and let $p \in R$ be irreducible.*

1. *For any prime q that is an associate of p, we have $ord_p(q) = 1$.*
2. *For any prime q that is not an associate of p, we have $ord_p(q) = 0$.*
3. *For any unit u, we have $ord_p(u) = 0$.*

Proof. 1. Suppose that q is a prime that is an associate of p. Fix a unit u with $q = pu$. Notice that if $p \mid u$, then since $u \mid 1$, we conclude that $p \mid 1$, which would imply that p is a unit. Since p is not a unit, it follows that $p \nmid u$. Therefore, $ord_p(q) = 1$ by Lemma 3.3.12.

2. Suppose that q is a prime that is not an associate of p. Since q is prime, it is irreducible, so its only divisors are units and associates. Since p is not a unit nor an associate of q, it follows that $p \nmid q$. Therefore, $ord_p(q) = 0$.

3. This is immediate because if $p \mid u$, then since $u \mid 1$, we could conclude that $p \mid 1$. This implies that p is a unit, which is a contradiction.

□

Lemma 3.3.15. *Let R be a PID. Let $a \in R$ and let $p \in R$ be irreducible. Suppose that u is a unit, that q_i are irreducibles, and that*

$$a = uq_1q_2 \cdots q_k$$

We then have that exactly $ord_p(a)$ many of the q_i are associates of p.

Proof. Since

$$a = uq_1q_2 \cdots q_k$$

we have

$$\begin{aligned} ord_p(a) &= ord_p(uq_1q_2 \cdots q_k) \\ &= ord_p(u) + \sum_{i=1}^{k} ord_p(q_i) \\ &= \sum_{i=1}^{k} ord_p(q_i) \end{aligned}$$

The terms on the right are 1 when q_i is an associate of p and 0 otherwise. The result follows. □

Definition 3.3.16. *A* Unique Factorization Domain*, or UFD, is an integral domain R such that:*

1. *Every nonzero nonunit is a product of irreducible elements.*
2. *If $q_1q_2 \cdots q_n = r_1r_2 \ldots r_m$ where each q_i and r_j are irreducible, then $m = n$ and there exists a permutation $\sigma \in S_n$ such that p_i and $q_{\sigma(i)}$ are associates for every i.*

Theorem 3.3.17. *Every PID is a UFD. Moreover, if R is a PID and*

$$uq_1q_2 \cdots q_\ell = wr_1r_2 \cdots r_\ell$$

where u and w are units, and each of the q_i and r_j are irreducible, then $k = \ell$ and there exists $\sigma \in S_k$ such that q_i and $r_{\sigma(i)}$ are associates for all i.

Proof. Let R be a PID. We know from above that every PID is Noetherian, and hence every nonzero nonunit in R is a product of irreducibles. Suppose now that

$$uq_1q_2 \cdots q_\ell = wr_1r_2 \cdots r_\ell$$

where u and w are units, and each of the q_i and r_j are irreducible. Let p be an arbitrary prime. We know from the lemma that exactly $ord_p(n)$ many of the q_i are associates of p, and also that exactly $ord_p(n)$ many of the r_j are associates of p. Thus, for every prime p, there are an equal number of associates of p on each side. Matching up the elements on the left with corresponding associates on the right gives the required permutation. □

3.4 Factorizations in the Gaussian Integers and Sums of Squares

Proposition 3.4.1. *An element $\alpha \in \mathbb{Z}[i]$ is a unit if and only if $N(\alpha) = 1$. Therefore, $U(\mathbb{Z}[i]) = \{1, -1, i, -i\}$,*

Proof. Suppose that $\alpha \in \mathbb{Z}[i]$ is a unit. Fix $\beta \in \mathbb{Z}[i]$ with $\alpha\beta = 1$. We then have

$$1 = N(1) = N(\alpha\beta) = N(\alpha)N(\beta)$$

Now $N(\alpha)$ and $N(\beta)$ are both nonnegative integers, so $N(\alpha) \mid 1$ in the integers, Therefore, $N(\alpha) = 1$,

Suppose conversely that $\alpha \in \mathbb{Z}[i]$ satisfies $N(\alpha) = 1$. Write $\alpha = a + bi$ where $a, b \in \mathbb{Z}$, We then have

$$1 = N(\alpha) = N(a + bi) = a^2 + b^2 = (a + bi)(a - bi)$$

Since $a - bi \in \mathbb{Z}[i]$, it follows that $a - bi$ has an inverse so is a unit.

Finally, to find all the units, we need only find all $\alpha \in \mathbb{Z}[i]$ with $N(\alpha) = 1$. That is, we need only find those pairs $a, b \in \mathbb{Z}$ so that

$$a^2 + b^2 = N(a + bi) = 1$$

If either $|a| \geq 2$ or $|b| \geq 2$, then $a^2 + b^2 \geq 4 > 1$. If both $|a| = 1 = |b|$, then $a^2 + b^2 = 1 + 1 = 2 > 1$. If both $|a| = 0 = |b|$, then $a^2 + b^2 = 0 + 0 = 0 < 1$. Thus, we must have exactly one of a or b in the set $\{1, -1\}$ and the other equal to the 0. This gives the four units $\{1, -1, i, -i\}$, □

Now 5 is of course prime (irreducible) in $\mathbb{Z}$. However, we have

$$5 = (2 + i)(2 - i)$$

and neither $2 + i$ nor $2 - i$ is a unit, so 5 is not irreducible in $\mathbb{Z}[i]$

The case of 3 is different. Of course, 3 is prime (irreducible) in $\mathbb{Z}$, but we just saw above that there is no reason to believe a priori that 3 is prime (irreducible) in $\mathbb{Z}[i]$. However, we now check that this is indeed the case. Suppose that $\alpha, \beta \in \mathbb{Z}[i]$ with $3 = \alpha\beta$. We then have that

$$9 = N(3) = N(\alpha\beta) = N(\alpha)N(\beta)$$

Therefore, we must have that $N(\alpha)$ and $N(\beta)$ are natural numbers which divide 9 in the integers. Thus, $N(\alpha) \in \{1, 3, 9\}$. However, we can not have $N(\alpha) = 3$ because there are no solutions to $a^2 + b^2 = 3$ in the integers. It follows that $N(\alpha) \in \{1, 9\}$. If $N(\alpha) = 1$, then α is a unit by the previous Proposition. If $N(\alpha) = 9$, then we must have $N(\beta) = 1$, so β is a unit by the previous Proposition. Therefore, whenever $\alpha, \beta \in \mathbb{Z}[i]$ and $3 = \alpha\beta$, then either α is a unit or β is a unit. It follows that 3 is irreducible (and hence prime) in $\mathbb{Z}[i]$.

Before examining the case of 5 further, we prove a simple lemma.

Lemma 3.4.2. *If $\alpha \in \mathbb{Z}[i]$ and $N(\alpha)$ is a prime in $\mathbb{Z}$, then α is prime in $\mathbb{Z}[i]$.*

Proof. Suppose that $\alpha = \beta\gamma$ in $\mathbb{Z}[i]$. We then have that

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

Since this is a product in $\mathbb{Z}$ and $N(\alpha)$ is prime in $\mathbb{Z}$, it follows that either $N(\beta) = 1$ or $N(\gamma) = 1$, so either β is a unit or γ is a unit. □

We saw above that $5 = (2 + i)(2 - i)$. The question arises as to whether the factors on the right break down further. Notice that

$$N(2 + i) = 2^2 + 1^2 = 5$$

and

$$N(2-i) = 2^2 + (-1)^2 = 5$$

Therefore, by the lemma, both of these factors are irreducible in $\mathbb{Z}[i]$. Therefore, the factorization

$$5 = (2+i)(2-i)$$

is the unique prime factorization in the ring $\mathbb{Z}[i]$. Notice that we also have

$$5 = (1+2i)(1-2i)$$

and both $1+2i$ and $1-2i$ are irreducible in $\mathbb{Z}[i]$ since they have norm 5. However, we have $1+2i = i(2-i)$ and $1-2i = -i(2-i)$, so the elements of this factorization are indeed associates of the one above,

Theorem 3.4.3. *Suppose that p is prime element of $\mathbb{Z}$. We then have that p can be written as the sum of two squares in $\mathbb{Z}$ if and only if p is no longer prime in $\mathbb{Z}[i]$*

Proof. Suppose first there exist $a, b \in \mathbb{Z}$ with $p = a^2 + b^2$. Notice that both $a \neq 0$ and $b \neq 0$ because p is not a square is $\mathbb{Z}$. In the ring $\mathbb{Z}[i]$, we then have that

$$p = a^2 + b^2 = (a+bi)(a-bi)$$

Now neither $a+bi$ nor $a-bi$ is a unit because $a \neq 0$ and $b \neq 0$ from above (so neither of them is $1, -1, i, -i$). Therefore, p is not irreducible and hence not prime in $\mathbb{Z}[i]$.

Suppose conversely that p is no longer prime in $\mathbb{Z}[i]$. We then have that p is not irreducible in $\mathbb{Z}[i]$, so there exist nonunits $\alpha, \beta \in \mathbb{Z}[i]$ with $p = \alpha\beta$. We then have

$$p^2 = N(p) = N(\alpha\beta) = N(\alpha)N(\beta)$$

Since $N(\alpha)$ and $N(\beta)$ are both natural numbers, we have that they are both in the set $\{1, p, p^2\}$. However, both are nonunits, so we know that $N(\alpha) \neq 1$ and $N(\beta) \neq 1$. It follows that $N(\alpha) = p = N(\beta)$, Writing $\alpha = a + bi$, we have

$$p = N(\alpha) = N(a+bi) = a^2 + b^2$$

so p is the sum of two squares in $\mathbb{Z}$. □

Theorem 3.4.4. *An odd prime $p \in \mathbb{Z}$ is the sum of two squares in $\mathbb{Z}$ if and only if $p \equiv 1 \pmod 4$.*

Proof. We saw at the very beginning of the course that every square is congruent to either 0 or 1 modulo 4. Thus, the sum of two squares is one of 0, 1, or 2 modulo 4. Since an odd prime p must satisfy either $p \equiv 1 \pmod 4$ or $p \equiv 3 \pmod 4$, it follows that if an odd prime is the sum of two squares then $p \equiv 1 \pmod 4$.

Suppose conversely that $p \equiv 1 \pmod 4$. By the previous theorem it suffices to show that p is not prime in $\mathbb{Z}[i]$. Since $p \equiv 1 \pmod 4$, we know that -1 is a square modulo p, so there exists $a \in \mathbb{Z}$ with $a^2 \equiv -1 \pmod p$. We then have that $p \mid (a^2+1)$ in $\mathbb{Z}$, so $p \mid (a^2+1)$ in $\mathbb{Z}[i]$ also because $\mathbb{Z}$ is a subring of $\mathbb{Z}[i]$. Now $a^2 + 1 = (a+i)(a-i)$, hence $p \mid (a+i)(a-i)$. However, in $\mathbb{Z}[i]$ we have both $p \nmid (a+i)$ and $p \nmid (a-i)$ (because $p \cdot (k + \ell i) = pk + (p\ell)i$ so any multiple of p in $\mathbb{Z}[i]$ must satisfy that both the real and imaginary parts are divisible by p in $\mathbb{Z}$). It follows that p is not prime in $\mathbb{Z}[i]$. Therefore, by the previous proposition, p is the sum of two squares in $\mathbb{Z}$. □

We summarize the results in the following theorem.

Theorem 3.4.5. *Let $p \in \mathbb{Z}$ be an odd prime. The following are equivalent.*

1. *p is the sum of two squares in $\mathbb{Z}$, i.e. there exist $a, b \in \mathbb{Z}$ with $p = a^2 + b^2$.*

2. *p is reducible (and hence no longer prime) in $\mathbb{Z}[i]$.*

3. *-1 is a square modulo p.*

4. *$p \equiv 1 \pmod 4$.*

Proof. We have $1 \leftrightarrow 2$ by Theorem 3.4.3, $3 \leftrightarrow 4$ by Theorem 2.8.1, and $1 \leftrightarrow 4$ by Theorem 3.4.4. However, we show how to prove $1 \to 2 \to 3 \to 2 \to 1$ directly to see how we could get by even if we did not have a classification for when -1 is a square modulo 4.

- $1 \to 2$: Suppose that $p = a^2 + b^2$. Notice that $a \neq 0$ and $b \neq 0$ because primes are not squares. We then have
$$p = (a + bi)(a - bi)$$
Since both $a \neq 0$ and $b \neq 0$, neither $a + bi$ nor $a - bi$ is a unit (because $U(\mathbb{Z}[i]) = \{\pm 1, \pm i\}$). Thus, p is reducible in $\mathbb{Z}[i]$.

- $2 \to 3$: Suppose first that $p \in \mathbb{Z}$ is an odd prime which is reducible in R. Fix $\alpha, \beta \in R$ both nonunits with $p = \alpha\beta$. Taking norms we see that
$$p^2 = N(p) = N(\alpha\beta) = N(\alpha)N(\beta)$$
Since the norm of every element of R is a nonnegative integer and both $N(\alpha) \neq 1$ and $N(\beta) \neq 1$ (because they are nonunits), it follows that $N(\alpha) = p = N(\beta)$. Let $\alpha = a + bi$ where $a, b \in \mathbb{Z}$. We then have
$$p = N(\alpha) = N(a + bi) = a^2 + b^2$$
Suppose that $p \mid b$ in $\mathbb{Z}$. We then have $p \mid (p - b^2)$ in $\mathbb{Z}$, so $p \mid a^2$ in $\mathbb{Z}$, and hence $p \mid a$ in $\mathbb{Z}$ because p is prime in $\mathbb{Z}$. Since $p \mid a$ and $p \mid b$ in $\mathbb{Z}$, we then have that $p^2 \mid a^2$ and $p^2 \mid b^2$ in $\mathbb{Z}$, so we could conclude that $p^2 \mid (a^2 + b^2)$ in $\mathbb{Z}$ and hence $p^2 \mid p$ in $\mathbb{Z}$ which is a contradiction. Therefore, $p \nmid b$ in $\mathbb{Z}$.
Now we know that $p = a^2 + b^2$, so $p \mid (a^2 - (-b^2))$ and hence
$$a^2 \equiv -b^2 \pmod p$$
We just showed that $p \nmid b$ in $\mathbb{Z}$, so $\gcd(b, p) = 1$, and hence we may fix $c \in \mathbb{Z}$ with $bc \equiv 1 \pmod p$. Multiplying both sides of the above congruence by c^2, we conclude that
$$a^2c^2 \equiv -b^2c^2 \pmod p$$
so $(ac)^2 \equiv -(bc)^2 \equiv -1 \pmod p$. Therefore, -1 is a square modulo p.

- $3 \to 1$: Fix $a \in \mathbb{Z}$ with $a^2 \equiv -1 \pmod p$. We then have that $p \mid (a^2 + 1)$ in $\mathbb{Z}$, so since $(a^2 + 1) = (a + i)(a - i)$ in $\mathbb{Z}[i]$, it follows that
$$p \mid (a + i)(a - i)$$
in $\mathbb{Z}[i]$. Now notice that $p(k + \ell i) = pk + p\ell i$ for any $k, \ell \in \mathbb{Z}$, so if $p \mid (c + di)$ in $\mathbb{Z}[i]$, then both $p \mid c$ and $p \mid d$ in $\mathbb{Z}$. Since $p \nmid 1$ and $p \nmid -1$ in $\mathbb{Z}$, it follows that $p \nmid (a + i)$ and $p \nmid (a - i)$ in $\mathbb{Z}[i]$. Therefore, we know that p is not prime in $\mathbb{Z}[i]$. Since $\mathbb{Z}[i]$ is a PID, it follows that p is reducible in $\mathbb{Z}[i]$ (notice that p is not a unit because its norm is not 1).

- $2 \to 1$: Suppose that p is reducible in $\mathbb{Z}[i]$. Fix nonunits $\alpha, \beta \in \mathbb{Z}[i]$ such that $p = \alpha\beta$. We then have that $N(p) = N(\alpha\beta)$, so $p^2 = N(\alpha) \cdot N(\beta)$. Now $N(\alpha) \neq 1$ and $N(\beta) \neq 1$ since they are not units, so as $N(\alpha)$ and $N(\beta)$ are nonnegative, we must have $N(\alpha) = p = N(\beta)$. Fixing $a, b \in \mathbb{Z}$ with $\alpha = a + bi$, we have
$$p = N(\alpha) = a^2 + b^2$$
which completes the proof.

□

Proposition 3.4.6. *If $p \equiv 1 \pmod 4$ is an odd prime, then there is a unique way to write p as the sum of two squares up to change in order and sign. In other words, if $a^2 + b^2 = p = c^2 + d^2$, then one of the following is true:*

- *$a = \pm c$ and $b = \pm d$*
- *$a = \pm d$ and $b = \pm c$*

Proof. Since $p \equiv 1 \pmod 4$, there exist $a, b \in \mathbb{Z}$ with $p = a^2 + b^2$ by the previous theorem. Suppose that $a, b, c, d \in \mathbb{Z}$ and that both $a^2 + b^2 = p = c^2 + d^2$. Now none of a, b, c, d is zero because p is not a square in $\mathbb{Z}$, so $N(a + bi) = a^2 + b^2 \geq 2$ and $N(c + di) = c^2 + d^2 \geq 2$. We have

$$N(a + bi) = N(a - bi) = p$$

so both $a + bi$ and $a - bi$ are irreducible in $\mathbb{Z}[i]$ by Lemma 3.4.2. Similarly we have that

$$N(c + di) = N(c - di) = p$$

so both $c + di$ and $c - di$ are irreducible in $\mathbb{Z}[i]$. Since $\mathbb{Z}[i]$ is a UFD and we have

$$(a + bi)(a - bi) = (c + di)(c - di)$$

where all factors are irreducible, we know that the factors on the left pair up with factors on the right as associates. In particular, either $a + bi$ and $c + di$ are associates, or $a + bi$ and $c - di$ are associates. There are four units in $\mathbb{Z}[i]$, namely $\{1, -1, i, -i\}$, so we now have 8 possible cases. Working through the 8 cases gives the above 8 possibilities. For example, if $a + bi = c + di$, then $a = c$ and $b = d$, while if $a + bi = (-i)(c - di)$, then $a + bi = -d - ci$ so $a = -d$ and $b = -c$. □

Now we move from primes to arbitrary integers. Clearly if an integer is the sum of two squares, then it is nonnegative. We also have the following.

Proposition 3.4.7. *For any $a, b, c, d \in \mathbb{Z}$, we have*

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$$

Proof. One proof is direct computation. We ahve

$$\begin{aligned}(a^2 + b^2) \cdot (c^2 + d^2) &= a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 \\ &= a^2c^2 - 2acbd + b^2d^2 + a^2d^2 + 2adbc + b^2c^2 \\ &= (ac - bd)^2 + (ad + bc)^2\end{aligned}$$

Alternatively, one can work in the Gaussian Integers to "see" the formula arise more naturally:

$$\begin{aligned}(a^2 + b^2)(c^2 + d^2) &= (a + bi)(a - bi)(c + di)(c - di) \\ &= (a + bi)(c + di)(a - bi)(c - di) \\ &= ((ac - bd) + (ad + bc)i) \cdot ((ac - bd) - (ad + bc)i) \\ &= (ac - bd)^2 + (ad + bc)^2\end{aligned}$$

□

Corollary 3.4.8. *If each of $m, n \in \mathbb{Z}$ can be written as the sum of two squares, then mn can also be written as the sum of two squares*

Proof. Immediate. □

Now we know that every prime $p \equiv 1 \pmod 4$ is the sum of two squares. Clearly $2 = 1^2 + 1^2$ is the sum of two squares, and $q^2 = q^2 + 0^2$ is trivially the sum of two squares for any prime $q \equiv 3 \pmod 4$. Thus, using the previous corollary, if $n \geq 2$ and every prime $q \equiv 3 \pmod 4$ appearing in the prime factorization of n occurs to an even power, then n is the sum of two squares. We now go about proving the converse.

Lemma 3.4.9. *Suppose that $a, b \in \mathbb{Z}$ and $q \in \mathbb{Z}$ is prime with $q \mid (a^2 + b^2)$. If $q \equiv 3 \pmod 4$, then $q \mid a$ and $q \mid b$.*

Proof. Let q be prime with $q \mid (a^2 + b^2)$. We prove the contrapositive. Suppose that either $q \nmid a$ or $q \nmid b$. Since $a^2 + b^2 = b^2 + a^2$, we can assume (by interchanging the roles of a and b if necessary) that $q \nmid b$. Since q is prime, it follows that $\gcd(b, q) = 1$. Thus, we may fix $c \in \mathbb{Z}$ with $bc \equiv 1 \pmod q$. Now $a^2 \equiv -b^2 \pmod q$, so multiplying both sides by c^2 we see that $(ac)^2 \equiv -1 \pmod q$. Thus, -1 is a square modulo q and hence $q \equiv 1 \pmod 4$. □

Theorem 3.4.10. *An integer $n \in \mathbb{Z}$ is the sum of two squares in $\mathbb{Z}$ if and only if either*

- *$n = 0$ or $n = 1$*
- *$n \geq 2$ and $ord_q(n)$ is even for all primes $q \equiv 3 \pmod 4$.*

Proof. We have $0 = 0^2 + 0^2$, $1 = 1^2 + 0^2$, and the existence of all of the others follows from the argument above. We prove that these are the only possibilities by induction on n. The base case is trivial. Suppose that $n \geq 2$ and that all smaller numbers satisfy the result. Suppose that $n = a^2 + b^2$. Let q be a prime with $q \equiv 3 \pmod 4$. If $q \nmid n$, then trivially $ord_q(n) = 0$ is even. Suppose then that $q \mid n$. By the lemma, we know that $q \mid a$ and $q \mid b$ so we may fix $k, \ell \in \mathbb{Z}$ with $a = qk$ and $b = q\ell$. We then have

$$n = a^2 + b^2 = (qk)^2 + (q\ell)^2 = q^2(k^2 + \ell^2)$$

Now $k^2 + \ell^2$ is a sum of two squares with $k^2 + \ell^2 < n$, so by induction we know that $ord_q(k^2 + \ell^2)$ is even. Therefore

$$ord_q(n) = 2 + ord_q(k^2 + \ell^2)$$

is also even. Since q was arbitrary prime, we conclude that $ord_q(n)$ is even for primes $q \equiv 3 \pmod 4$. This complete the induction. □

Theorem 3.4.11. *Up to associates, the prime elements of $\mathbb{Z}[i]$ are the following:*

- *$1 + i$*
- *Every prime number $q \in \mathbb{Z}$ which satisfies $q \equiv 3 \pmod 4$.*
- *For every prime number $p \in \mathbb{Z}$ which satisfies $p \equiv 1 \pmod 4$, if we write $p = a^2 + b^2$ for the unique choice of a and b satisfying $1 \leq a < b$, the elements $a + bi$ and $a - bi$.*

Proof. We first show that each of the above are indeed prime elements of $\mathbb{Z}[i]$. Notice that $N(1 + i) = 1^2 + 1^2 = 2$ is prime in $\mathbb{Z}$, so $1 + i$ is prime in $\mathbb{Z}[i]$ by Lemma 3.4.2. Also, if $p \equiv 1 \pmod 4$ and we have $p = a^2 + b^2$, then $N(a + bi) = N(a - bi) = p$ a prime in $\mathbb{Z}$, so again $a + bi$ and $a - bi$ are prime in $\mathbb{Z}[i]$ by Lemma 3.4.2. Suppose that q is prime in $\mathbb{Z}$ and that $q \equiv 3 \pmod 4$. Since q is not the sum of two squares by Theorem 3.4.4, we know that q is remains prime in $\mathbb{Z}[i]$ by Proposition 3.4.3. Thus, each of the above are prime. We need only show that none of the above are associates. Now associates have the same norm, and we have $N(1 + i) = 2$ and $N(q) = q^2$ for all $q \equiv 3 \pmod 4$, so none of these are associates. Suppose that $p \equiv 1 \pmod 4$ and we write $p = a^2 + b^2$ for the unique choice of a and b satisfying $1 \leq a < b$. We have $N(a + bi) = p = N(a - bi)$, so these elements are not associates of any others. Finally, a simple check

through the four units shows that $a+bi$ and $a-bi$ are not associates of each other either (here we make essential use of the the fact that $a \neq b$ because $p = a^2+b^2$ is an odd prime and hence $p \neq 2a^2$ for any $a \in \mathbb{Z}$).

Suppose that $\pi \in \mathbb{Z}[i]$ is prime. Let $n = N(\pi) \in \mathbb{N}^+$. If $n = 0$, then $\pi = 0$ so π is not prime. Also, if $n = 1$, then π is a unit so π is not prime. Suppose then $n \geq 2$. Write n as a product of primes in $\mathbb{Z}$ as $n = p_1 p_2 \cdots p_k$. Notice that $\pi\overline{\pi} = N(\pi) = n$ (where $\overline{\pi}$ is the complex conjugate of π), so $\pi \mid n$ in $\mathbb{Z}[i]$. Hence, in $\mathbb{Z}[i]$ we have that

$$\pi \mid p_1 p_2 \cdots p_k$$

and since π is prime it follows that π divides some element on the right in $\mathbb{Z}[i]$. In particular, we may fix a prime number $p \in \mathbb{Z}$ such that $\pi \mid p$ in $\mathbb{Z}[i]$. We know have three cases.

- Suppose that $p \equiv 3 \pmod 4$. Since p is prime (irreducible) in $\mathbb{Z}[i]$ and $\pi \mid p$, we know that either π is a unit or π is an associate of p. The former is impossible, so π is an associate of p.
- Suppose that $p \equiv 1 \pmod 4$. Writing $p = a^2 + b^2$ where $1 \leq a < b$, we have that $p = (a+bi)(a-bi)$ and each of these factors are prime (irreducible) in $\mathbb{Z}[i]$ because they have norm p (which is prime in $\mathbb{Z}$). Since π is prime in $\mathbb{Z}[i]$ and $\pi \mid (a+bi)(a-bi)$, we have that π divides one of these factors, and since they are both irreducible and π is not a unit, it follows that π is an associate of one of these factors.
- Suppose that $p = 2$. Now $2 = (1+i)(1-i)$ and each of these factors are prime (irreducible) in $\mathbb{Z}[i]$ because they have norm 2. Since π is prime in $\mathbb{Z}[i]$ and $\pi \mid (1+i)(1-i)$, either $\pi \mid (1+i)$ and $\pi \mid (1-i)$. Thus, π is an associate of either $1+i$ or $1-i$. Finally, notice that $(1-i) = -i(1+i)$, so these two factors are associates of each other, and thus in either case π is an associate of $1+i$.

□

3.5 Pythagorean Triples and Diophantine Equations

Theorem 3.5.1. *Let R be a PID. Suppose that $r, a, b \in R$ and $n \in \mathbb{N}^+$ is such that $r^n = ab$. Suppose also that a and b are relatively prime. There exists $u \in U(R)$ and $s \in R$ with $a = us^n$. Similarly, there exists $v \in U(R)$ and $t \in R$ with $b = vt^n$.*

Proof. Let $p \in R$ be an arbitrary irreducible. Applying ord_p to both sides of $r^n = ab$ gives

$$n \cdot ord_p(r) = ord_p(a) + ord_p(b)$$

Notice that we can not have $ord_p(a)$ and $ord_p(b)$ both positive (since this would imply p is a nonunit common divisor of a and b). Now if $ord_p(a) = 0$, then $n \mid ord_p(b)$ in $\mathbb{Z}$, while if $ord_p(b) = 0$, then $n \mid ord_p(a)$ in $\mathbb{Z}$. Therefore, for every irreducible $p \in R$, we have both $n \mid ord_p(a)$ in $\mathbb{Z}$ and $n \mid ord_p(b)$ in $\mathbb{Z}$.

Now write a as a product of irreducibles

$$a = q_1 q_2 \cdots q_\ell$$

Since $n \mid ord_p(a)$ for all irreducibles $p \in R$, it follows that the number of associates of p in the above product is a multiple of n. Factoring out units to make these associates equal, it follows that we can write

$$a = u p_1^{nk_1} p_2^{nk_2} \cdots p_\ell^{nk_\ell}$$

where u is a unit and the p_i are not associates of each other. We then have

$$a = u \cdot (p_1^{k_1} p_2^{k_2} \cdots p_\ell^{k_\ell})^n$$

The argument for b is completely analogous. □

3.5.1 Pythagorean Triples

Using this theorem, one can solve certain Diophantine equations by working over a PID such as $\mathbb{Z}[i]$. We first show how to work in $\mathbb{Z}[i]$ to derive parameterizations of the primitive Pythagorean triples as we did in Theorem 1.7.6. This approach is slightly faster than the elementary one we presented, and it has the key advantage of "explaining" the formulas as arising from multiplication in $\mathbb{C}$.

Assume then that (a, b, c) is a primitive Pythagorean triple. Recall that we know that exactly one of a and b is even and that c is odd from Proposition 1.7.5. Now we have

$$c^2 = a^2 + b^2 = (a + bi)(a - bi)$$

We claim that $a + bi$ and $a - bi$ are relatively prime in $\mathbb{Z}[i]$. In order to show this, it suffices to show that $a + bi$ and $a - bi$ have no irreducible common divisor. Suppose then that δ is an irreducible common divisor of $a + bi$ and $a - bi$. We then have that δ divides both

$$(a + bi) + (a - bi) = 2a \qquad \text{and} \qquad (a + bi) - (a - bi) = 2bi$$

Notice that since $\delta \mid 2bi$, we also know that δ divides $2bi \cdot (-i) = 2b$. Thus, δ is a common divisor of $2a$ and $2b$ in $\mathbb{Z}[i]$. Now in $\mathbb{Z}$, we have that a and b are relatively prime, so there exists $m, n \in \mathbb{Z}$ with $ma + nb = 1$. We then have $m \cdot 2a + n \cdot 2b = 2$, so $\delta \mid 2$ in $\mathbb{Z}[i]$. Recall that in $\mathbb{Z}[i]$, we have

$$2 = (-i)(1 + i)^2$$

Thus, since δ is irreducible, we must have that δ is an associate of $1 + i$. In particular, we conclude that $1 + i \mid \delta$. By taking norms we infer that $2 \mid N(\delta)$ in $\mathbb{Z}$. Since $\delta \mid a + bi$, this implies that $\delta \mid c^2$ in $\mathbb{Z}[i]$, and hence $N(\delta) \mid c^4$ in $\mathbb{Z}$. Therefore, we would have that c is even, which is a contradiction.

We now have

$$c^2 = (a + bi)(a - bi)$$

where $a + bi$ and $a - bi$ are relatively prime in $\mathbb{Z}[i]$. Using the above theorem, we may fix $\mu \in U(\mathbb{Z}[i])$ and $\gamma \in \mathbb{Z}[i]$ with

$$a + bi = \mu \cdot \gamma^2$$

Notice that $-1 = i^2$ is a square in $\mathbb{Z}[i]$, so it can be absorbed into γ and hence there exists $\gamma \in \mathbb{Z}[i]$ such that either $a + bi = \gamma^2$ or $a + bi = i \cdot \gamma^2$. Fix $m, n \in \mathbb{Z}$ with $\gamma = n + mi$. We then have

$$\gamma^2 = (n + mi)^2 = [(n^2 - m^2) + 2mn \cdot i]$$

We now have two cases.

- If $a + bi = \gamma^2$, then $a = n^2 - m^2$ and $b = 2mn$.
- If $a + bi = i\gamma^2$, then $a = -2mn$ and $b = n^2 - m^2$.

Suppose that we also assume that $a, b \in \mathbb{N}^+$ (we require this for primitive Pythagorean triples, but had not yet used it to this point). In the first case we can assume that $m, n \in \mathbb{N}^+$ (by replacing them by their negatives if both are negative) and that $m < n$. In the second case, we must have that m and n have different parities. By replacing the negative value with its absolute value, we obtain $a = 2mn$ and $b = n^2 - m^2$ where both $m, n \in \mathbb{N}^+$ and $m < n$. From here, it is easy to show that $c = m^2 + n^2$ in both cases.

3.5.2 Squares and Cubes

Suppose that we try to find all integer solutions to

$$x^3 = y^2 + 1$$

The solution $(1, 0)$ is clear, but are there any others? Notice that if y is odd, then $y^2 + 1 \equiv 2 \pmod 4$, but 2 is not a cube modulo 4. Thus, we can assume that y is even and hence that x is odd. We can factor the right-hand side as

$$x^3 = (y + i)(y - i)$$

We claim that $y + i$ and $y - i$ are relatively prime in $\mathbb{Z}[i]$. As above, it suffices to show that $y + i$ and $y - i$ have no irreducible common divisor. Suppose then that $\delta \in \mathbb{Z}[i]$ is an irreducible common divisor of $y + i$ and $y - i$. We then have that δ divides

$$(y + i) - (y - i) = 2i$$

and hence $\delta \mid 2$ in $\mathbb{Z}[i]$. Since $2 = (-i)(1 + i)^2$, it follows that we must have δ is a unit multiple of $1 + i$, so in particular $1 + i \mid \delta$. This implies that $(1 + i) \mid x^3$ in $\mathbb{Z}[i]$. Taking norms, we conclude that $2 \mid x^6$ in $\mathbb{Z}$, so x is even, a contradiction. Therefore, we have that $y + i$ and $y - i$ are relatively prime elements of $\mathbb{Z}[i]$ such that the product is a cube. By our theorem, there is a unit μ and a $\gamma \in \mathbb{Z}[i]$ with

$$y + i = \mu \cdot \gamma^3$$

Notice that all of the units of $\mathbb{Z}[i]$ are cubes, so there exists $\gamma \in \mathbb{Z}[i]$ with $y + i = \gamma^3$. Fixing $m, n \in \mathbb{Z}$ with $\gamma = m + ni$, we then have

$$y + i = (m + ni)^3 = (m^3 - 3mn^2) + (3m^2n - n^3)i$$

so in $\mathbb{Z}$ we have the equations

$$y = m(m^2 - 3n^2) \qquad \text{and} \qquad 1 = n(3m^2 - n^2)$$

The right-hand equation implies that $n = \pm 1$. If $n = 1$, we get $1 = 3m^2 - 1$, so $2 = 3m^2$, a contradiction. If $n = -1$, we get $1 = -(3m^2 - 1)$, so $-3m^2 = 0$ and hence $m = 0$. Plugging in these values we see that $y = 0$, and hence $x = 1$. Therefore, $(0, 1)$ is the only solution.

3.6 Ideals and Quotients of the Gaussian Integers

Suppose that I is a nonzero ideal of $\mathbb{Z}[i]$. Since $\mathbb{Z}[i]$ is a PID, we know that I is a principal ideal, and hence we may fix $\alpha \in \mathbb{Z}[i]$ with $\alpha \neq 0$ such that $I = \langle \alpha \rangle$, so

$$I = \langle \alpha \rangle = \{\gamma\alpha : \gamma \in \mathbb{Z}[i]\}$$

Now an arbitrary $\gamma \in \mathbb{Z}[i]$ can be written in the form $\gamma = c + di$ where $c, d \in \mathbb{Z}$, so

$$I = \{c\alpha + d \cdot i \cdot \alpha : c, d \in \mathbb{Z}\}$$

Viewed geometrically, this is saying that I is the *lattice* generated by the points α and $i \cdot \alpha$ in $\mathbb{Z}[i]$. Now $i \cdot \alpha$ is the result of rotating α counterclockwise by 90^o. Thus, we are taking two orthogonal vectors α and $i \cdot \alpha$ and forming all possible integers combinations of these elements. For example, if $\alpha = 3$, then we get the lattice generated by 3 and $3i$. On the other hand, if $\alpha = 2 + 2i$, then we the get the lattice generated by $2 + 2i$ and $-2 + 2i$, while if $\alpha = 1 + 2i$, then we get the lattice generated by $1 + 2i$ and $-2 + i$.

From the homework, you know that $\mathbb{Z}[i]/I$ is a finite ring, and in fact every element of the quotient ring equals $\rho + I$ for some $\rho \in \mathbb{Z}[i]$ with $N(\rho) < N(\alpha)$. You did not prove (and it is not true) that these ρ give distinct cosets. It is an interesting problem to determine the size of $\mathbb{Z}[i]/I$ based on a generator for I.

Proposition 3.6.1. *Let $n \in \mathbb{Z}$ with $n \neq 0$. We then have that $|\mathbb{Z}[i]/\langle n\rangle| = n^2$.*

Proof. Let

$$R = \{c + di \in \mathbb{Z}[i] : 0 \leq c < n \text{ and } 0 \leq d < n\}$$

Notice that $|R| = n^2$. We claim that the elements of R provide representatives for all of the cosets of I, and that no two elements of R are in the same coset of I. Let $\alpha \in \mathbb{Z}[i]$ be arbitrary, and write $\alpha = a + bi$ where $a, b \in \mathbb{Z}$. Using division in $\mathbb{Z}$, write

$$a = q_1 n + r_1 \qquad \text{and} \qquad b = nq_2 + r_2$$

where $q_i, r_i \in \mathbb{Z}$ and $0 \leq r_i < n$ (notice this is still possible even if a or b are negative). We then have

$$\begin{aligned}(a + bi) - (r_1 + r_2 i) &= (a - r_1) + (b - r_2)i \\ &= q_1 n + q_2 n \cdot i \\ &= n \cdot (q_1 + q_2 i)\end{aligned}$$

hence

$$\alpha - (r_1 + r_2 i) \in I$$

and therefore

$$\alpha + I = (r_1 + r_2 i) + I$$

Thus, the elements of R provide representatives for all of the cosets of I.

We now need to show that distinct representatives of R lie in different cosets of I. Suppose that

$$(c_1 + d_1 i) + I = (c_2 + d_2 i) + I$$

where $c_i, d_i \in \mathbb{Z}$, $0 \leq c_i < n$ and $0 \leq d_i < n$. We then have

$$(c_1 - c_2) + (d_1 - d_2)i \in I$$

so we may fix $a, b \in \mathbb{Z}$ with

$$n(a + bi) = (c_1 - c_2) + (d_1 - d_2)i$$

This implies that $na = c_1 - c_2$ and $nb = d_1 - d_2$. Therefore, $n \mid (c_1 - c_2)$ and $n \mid (d_1 - d_2)$. Since $-|n| < c_1 - c_2 < |n|$ and $-|n| < d_1 - d_2 < |n|$, it follows that $c_1 - c_2 = 0$ and $d_1 - d_2 = 0$. Therefore, $c_1 = d_1$ and $c_2 = d_2$. □

Corollary 3.6.2. *Let $p \in \mathbb{N}^+$ be prime with $p \equiv 3 \pmod 4$. We then have that $\mathbb{Z}[i]/\langle p\rangle$ is a field with p^2 elements.*

Proof. Since $p \equiv 3 \pmod 4$, we know that p is irreducible in $\mathbb{Z}[i]$. Since $\mathbb{Z}[i]$ is a PID, this implies that $\langle p\rangle$ is a maximal ideal of $\mathbb{Z}[i]$. Therefore, $\mathbb{Z}[i]/\langle p\rangle$ is a field with p^2 elements. □

Let's think about the case where $I = \langle\alpha\rangle$ but $\alpha \notin \mathbb{Z}$. If we draw the lattice corresponding to $\alpha = 1 + 2i$, then geometrically it appears that

$$0 \quad i \quad 2i \quad -1 + i \quad -1 + 2i$$

provide distinct representatives for the quotient $\mathbb{Z}[i]/\langle\alpha\rangle$. These were determined by picking points within one of the squares formed by the lattice $\langle\alpha\rangle$. For a general such α, it can tiresome to find such representatives and it may not be obvious how many there will be. On the homework, you will show that $\mathbb{Z}[i]/\langle\alpha\rangle$ has $N(\alpha)$ many elements.

Chapter 4

Field Extensions

4.1 Degree of an Extension

Notation 4.1.1. *Let F and K be fields. We often abuse notation and write $F \subseteq K$ to mean that F is a subfield of K (not merely a subset). We call $F \subseteq K$ a* field extension.

Since any intersection of subfields of K is itself a subfield of K, we can make the following definition.

Definition 4.1.2. *Let $F \subseteq K$ be a field extension and let $\alpha \in K$.*

- *The set $F[\alpha]$ is defined to be the smallest subring of K containing $F \cup \{\alpha\}$.*
- *The set $F(\alpha)$ is defined to be the smallest subfield of K containing $F \cup \{\alpha\}$.*

For example, working with the field extension $\mathbb{Q} \subseteq \mathbb{C}$, we have that $\mathbb{Q}[i] = \{a+bi : a, b \in \mathbb{Q}\}$ because it is straightforward to check that this set is closed under addition and multiplication. In fact, we showed above that this set is actually a field and it is clearly the smallest field containing $\mathbb{Q} \cup \{i\}$ because any field must be closed under addition and multiplication. We conclude that $\mathbb{Q}[i] = \mathbb{Q}(i)$. Similarly, it is straightforward to check that $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$, and that $\mathbb{Q}[2]$ is a field, so $\mathbb{Q}[\sqrt{2}] = \mathbb{Q}(\sqrt{2})$.

One issue that is not obvious in the description of $\mathbb{Q}(\sqrt{2})$ is the uniqueness of representation. Suppose that $a, b, c, d \in \mathbb{Q}$ with $a + b\sqrt{2} = c + d\sqrt{2}$. We then have that $(a - c) = (d - b)\sqrt{2}$. Now if $d - b \neq 0$, then $\sqrt{2} = \frac{a-c}{d-b}$ would be rational, a contradiction. Thus, we must have $d - b = 0$ and hence $b = d$. Canceling $b\sqrt{2} = d\sqrt{2}$ from both sides of $a + b\sqrt{2} = c + d\sqrt{2}$ we conclude that $a = c$.

The following proposition gives us a "constructive" way to determine $F[\alpha]$.

Proposition 4.1.3. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$. We have*

$$F[\alpha] = \{p(\alpha) : p(x) \in F[x]\}$$

Proof. Notice that if $p(x) \in F[x]$, then $p(\alpha)$ is simply a sum of products of elements of $F \cup \{\alpha\}$, hence we must have $p(\alpha) \in F[\alpha]$. It follows that $\{p(\alpha) : p(x) \in F[x]\} \subseteq F[\alpha]$. To show that converse containment, it suffices to show that $\{p(\alpha) : p(x) \in F[x]\}$ is in fact a subring of K containing $F \cup \{\alpha\}$. Notice that $0, 1 \in \{p(\alpha) : p(x) \in F[x]\}$ by considering the zero polynomial and one polynomial. Given two polynomials $p(x), q(x) \in F[x]$, we have $p(\alpha) + q(\alpha) = (p + q)(\alpha)$ and $p(\alpha) \cdot q(\alpha) = (pq)(\alpha)$. Since $F \cup \{\alpha\} \subseteq \{p(\alpha) : p(x) \in F[x]\}$ by considering the constant polynomials and $x \in F[x]$, it follows that $\{p(\alpha) : p(x) \in F[x]\}$ is indeed a subring of K containing $F \cup \{\alpha\}$. Thus, $F[x] \subseteq \{p(\alpha) : p(x) \in F[x]\}$. □

As we've seen in the cases of $\mathbb{Q}[i]$ and $\mathbb{Q}[\sqrt{2}]$, sometimes you can describe $F[\alpha]$ using far less than all polynomials since in both of those cases we can get away with linear polynomials. However, consider $\mathbb{Q}[\sqrt[3]{2}]$. A first guess might be that this ring equals

$$\{a + b\sqrt[3]{2} : a, b \in \mathbb{Q}\}$$

However, it's not clear that this is a subring of $\mathbb{R}$ because it's not obvious that $\sqrt[3]{4} = \sqrt[3]{2} \cdot \sqrt[3]{2}$ is in this set. In fact, it is not, but rather than work through the details now, we will develop general tools in the next section to figure out how to describe $F[\alpha]$ using a smaller collection of polynomials depending on the properties of α.

The following result can be proved similarly. Again, sometimes we can get away with far less (like in $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{2})$), but this is certainly enough.

Proposition 4.1.4. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$. We have*

$$F(\alpha) = \{\frac{p(\alpha)}{q(\alpha)} : p(x), q(x) \in F[x] \text{ and } q(\alpha) \neq 0\}$$

Suppose that $F \subseteq K$ is a field extension. We can view K as a vector space over the field F where vector addition is just addition in the field K, and scalar multiplication of a an element of F by an element of K is just multiplication in the field K. Notice that each of the vector space axioms hold by the properties of a field.

In the field extension $\mathbb{Q} \subseteq \mathbb{Q}(i)$, we have that $\{1, i\}$ is a basis for $\mathbb{Q}(i)$ over $\mathbb{Q}$ (simply because this set spans $\mathbb{Q}(i)$ over $\mathbb{Q}$ as described above, and its linearly independent because if $q + ri = 0$, then $q = 0$ and $r = 0$).

Definition 4.1.5. *Given a field extension $F \subseteq K$, we let $[K : F]$ be the dimension of K as a vector space over F (if there is no finite basis, we let $[K : F] = \infty$). The number $[K : F]$ is called the* degree *of K over F.*

For example, we have $[\mathbb{Q}(i) : \mathbb{Q}] = 2$ because $\{1, i\}$ is a basis over $\mathbb{Q}$ as discussed above. We also have $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$ because $\{1, \sqrt{2}\}$ spans $\mathbb{Q}(\sqrt{2})$ over $\mathbb{Q}$ and one aspect of the uniqueness of representation was that $\{1, \sqrt{2}\}$ is linearly independent over $\mathbb{Q}$.

4.2 Algebraic and Transcendental Elements

Definition 4.2.1. *Suppose that $F \subseteq K$ is a field extension. An element $\alpha \in K$ is* algebraic *over F if there exists a nonzero polynomial $p(x) \in F[x]$ such that $p(\alpha) = 0$. An element of K which is not algebraic over F is said to be* transcendental *over F.*

For example, $\sqrt{2}$ is algebraic over $\mathbb{Q}$ because it is a root of $x^2 - 2$. In fact, for every $n \in \mathbb{N}^+$, we have that $\sqrt[n]{2}$ is algebraic over $\mathbb{Q}$ since it is a root of $x^n - 2$. We have that i is algebraic over $\mathbb{Q}$ because it is a root of $x^2 + 1$. Every $q \in \mathbb{Q}$ is algebraic over $\mathbb{Q}$ since it is a root of $x - q$.

For a more interesting example, consider $\alpha = \sqrt{2} + \sqrt{3}$. We claim that α is algebraic over $\mathbb{Q}$. To find a polynomial with rational coefficients of which α is a root, we first calculate

$$\alpha^2 = (\sqrt{2} + \sqrt{3})^2 = 2 + 2 \cdot \sqrt{2} \cdot \sqrt{3} + 3 = 5 + 2\sqrt{6}$$

so

$$\alpha^2 - 5 = 2\sqrt{6}$$

Squaring both sides we conclude that

$$\alpha^4 - 10\alpha^2 + 25 = 24$$

and hence

$$\alpha^4 - 10\alpha^2 + 1 = 0$$

Therefore, $\alpha = \sqrt{2} + \sqrt{3}$ is a root of the polynomial $x^4 - 10x^2 + 1 \in \mathbb{Q}[x]$. In fact, we will see later through a more sophisticated argument that the sum, product, and quotient of algebraic elements is itself algebraic, i.e. the set of elements of K which are algebraic over F forms a subfield of K.

At this point, it is not obvious that there are elements of $\mathbb{C}$ (or even $\mathbb{R}$) which are transcendental over $\mathbb{Q}$. It is possible to prove that such elements exist by cardinality considerations, but it is also possible to prove that certain important analytic numbers are transcendental. In particular, both π and e are transcendental over $\mathbb{Q}$, though proving these requires some nontrivial analytic work. We will (unfortunately) not pursue that here.

Definition 4.2.2. *Suppose that $F \subseteq K$ is a field extension. Given $\alpha \in K$, we let*

$$I_\alpha = \{p(x) \in F[x] : p(\alpha) = 0\}$$

Proposition 4.2.3. *Suppose that $F \subseteq K$ is a field extension. For any $\alpha \in K$, the set I_α is a proper ideal of $F[x]$.*

Proof. Clearly $0 \in I_\alpha$. If $p(x), q(x) \in I_\alpha$, then $(p+q)(\alpha) = p(\alpha) + q(\alpha) = 0 + 0 = 0$, so $p(x) + q(x) \in I_\alpha$. If $p(x) \in I_\alpha$ and $f(x) \in F[x]$, then $(fp)(\alpha) = f(\alpha) \cdot p(\alpha) = f(\alpha) \cdot 0 = 0$, so $f(x) \cdot p(x) \in I_\alpha$. Combining this all, we conclude that I_α is an ideal of $F[x]$. Notice that $I_\alpha \neq F[x]$ because $1 \notin I_\alpha$. □

You can also prove the preceding proposition in the following way. Consider the ring homomorphism $Ev_\alpha \colon F[x] \to K$ given by $Ev_\alpha(p(x)) = p(\alpha)$. Notice that $ker(Ev_\alpha) = I_\alpha$, hence I_α is an ideal of $F[\alpha]$ (because kernels of ring homomorphisms are always ideals). Moreover, by Proposition 4.1.3, we have that $range(Ev_\alpha) = F[\alpha]$, so by the First Isomorphism Theorem, we conclude that

$$F[x]/I_\alpha \cong F[\alpha]$$

Notice that an element $\alpha \in K$ is algebraic over F if and only if $I_\alpha \neq \{0\}$. Recall that if F is a field, then $F[x]$ is a PID (because it is a Euclidean Domain). Hence, if $F \subseteq K$ is a field extension, then for every $\alpha \in K$, the ideal I_α is principal. If $\alpha \in K$ is algebraic over F, then $I_\alpha \neq \{0\}$, so any generator of I_α is nonzero. Recall that generators of an ideal are unique up to associates, so there is a unique monic generator of I_α.

Definition 4.2.4. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$ is algebraic over F. Since $F[x]$ is a PID, there exists a unique monic polynomial $m(x) \in F[x]$ with $I_\alpha = \langle m(x) \rangle$. The unique such monic polynomial is called the* minimal polynomial *of α over F.*

Since $I_\alpha = \langle m(x) \rangle$ where $m(x)$ is the minimal polynomials of α over F, we can rewrite the above ring isomorphism as:

$$F[x]/\langle m(x) \rangle \cong F[\alpha]$$

Proposition 4.2.5. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$ is algebraic over F. The minimal polynomial $m(x)$ of α over F is irreducible in $F[x]$.*

Proof. Let $m(x) \in F[x]$ be the minimal polynomial of α over F, so $I_\alpha = \langle m(x) \rangle$. Suppose that $p(x), q(x) \in F[x]$ with $m(x) = p(x)q(x)$. We then have

$$0 = m(\alpha) = p(\alpha) \cdot q(\alpha)$$

Since F is a field, it is an integral domain, and hence either $p(\alpha) = 0$ or $q(\alpha) = 0$. Suppose that $p(\alpha) = 0$. We then have $p(x) \in I_\alpha = \langle m(x) \rangle$, and hence $m(x) \mid p(x)$. Fixing $f(x) \in F[x]$ with $m(x)f(x) = p(x)$, we then have

$$m(x) = p(x)q(x) = m(x)f(x)q(x)$$

so as $F[x]$ is an integral domain, it follows that $1 = f(x)q(x)$ and so $q(x)$ is a unit. Similarly, if $q(\alpha) = 0$, then $p(x)$ is a unit. It follows that $m(x)$ is irreducible. □

Proposition 4.2.6. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$ is algebraic over F. If $p(x) \in F[x]$ is a monic irreducible polynomial with $p(\alpha) = 0$, then $p(x)$ is the minimal polynomial of α over F.*

Proof. Let $m(x) \in F[x]$ be the minimal polynomial of α over F. Since $p(\alpha) = 0$, we have that $p(x) \in I_\alpha = \langle m(x) \rangle$, so $m(x) \mid p(x)$ in $F[x]$. We are assuming that $p(x)$ is irreducible, so either $m(x)$ is a unit or $m(x)$ is an associate of $p(x)$. The former is impossible because I_α is a proper ideal of $F[x]$ (recall that $1 \notin I_\alpha$), hence $m(x)$ is an associate of $p(x)$. The units of $F[x]$ are just the constant polynomials, so as both $m(x)$ and $p(x)$ are monic, it follows that $p(x) = m(x)$. □

For example, the minimal polynomial of i over $\mathbb{Q}$ is $x^2 + 1$ (this polynomial does indeed have i as a root, and it is irreducible because it has degree 2 and no roots in $\mathbb{Q}$). Similarly, the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$. How about the minimal polynomial of $\sqrt[n]{2}$ over $\mathbb{Q}$ for some $n > 2$? Clearly, $\sqrt[n]{2}$ is a root of the polynomial $x^n - 2$. In fact, this polynomial is irreducible which can be seen by Eisenstein's Criterion, so the minimal polynomial of $\sqrt[n]{2}$ over $\mathbb{Q}$ is indeed $x^n - 2$.

Proposition 4.2.7. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$ is algebraic over F. Let $m(x) \in F[x]$ be the minimal polynomial of α over F, and let $n = \deg(m(x))$. We then have*

$$F[\alpha] = \{0\} \cup \{h(\alpha) : h(x) \in F[x] \text{ and } \deg(h(x)) < n\}$$

Proof. By Proposition 4.1.3, we know that $F[\alpha] = \{p(\alpha) : p(x) \in F[x]\}$, so clearly the set on the right is contained in $F[\alpha]$. Suppose now that $p(x) \in F[x]$ is arbitrary. Since $F[x]$ is a Euclidean Domain with Euclidean function equal to the degree map, there exists $q(x), r(x) \in F[x]$ with $p(x) = q(x)m(x) + r(x)$ and either $r(x) = 0$ or $\deg(r(x)) < \deg(m(x))$. We then have that

$$p(\alpha) = q(\alpha)m(\alpha) + r(\alpha) = q(\alpha) \cdot 0 + r(\alpha) = r(\alpha)$$

Thus, $p(\alpha) \in \{0\} \cup \{h(\alpha) : h(x) \in F[x] \text{ and } \deg(h(x)) < n\}$. It follows that $F[\alpha]$ is a subset of the set on the right. The result follows. □

Proposition 4.2.8. *If $F \subseteq K$ is a field extension and $\alpha \in K$ is algebraic over F, then $F(\alpha) = F[\alpha]$.*

Proof. We need only show that $F[\alpha]$ is a field. Let $m(x)$ be the minimal polynomial of α over F, and let $n = deg(m(x))$. We give two proofs.

The slick proof is to recall from above that

$$F[x]/\langle m(x) \rangle \cong F[\alpha]$$

Now $m(x)$ is irreducible by Proposition 4.2.5, so $\langle m(x) \rangle$ is a maximal ideal by Proposition 3.2.12. It follows that $F[x]/\langle m(x) \rangle$ is a field, so the isomorphic ring $F[\alpha]$ must be a field.

We now give a more constructive proof (which is really just unwrapping the "slick" proof above into the various pieces). Let $\beta \in F[\alpha]$ with $\beta \neq 0$. We need to show that $\beta^{-1} \in F[\alpha]$. We know that $F[\alpha] = \{0\} \cup \{h(\alpha) : h(x) \in F[x] \text{ and } \deg(h(x)) < n\}$, so we may fix a polynomial $h(x) \in F[x]$ with $\deg(h(x)) < n$ and $\beta = h(\alpha)$. Notice that $m(x) \nmid h(x)$ in $F[x]$ because $h(x) \neq 0$ and $\deg(h(x)) < n = \deg(m(x))$. Since $m(x)$ is irreducible, it follows that $\gcd(h(x), m(x)) = 1$. Fix polynomials $p(x), q(x) \in F[x]$ with

$$p(x)h(x) + q(x)m(x) = 1$$

We then have that

$$p(\alpha)h(\alpha) + q(\alpha)m(\alpha) = 1$$

so since $m(\alpha) = 0$, we conclude that $p(\alpha)h(\alpha) = 1$. Since $h(\alpha) = \beta$ and $p(\alpha) \in F[\alpha]$, we have shown that $\beta^{-1} = p(\alpha) \in F[\alpha]$. Therefore, $F[\alpha]$ is a field. □

Theorem 4.2.9. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$ is algebraic over F. Let $m(x) \in F[x]$ be the minimal polynomial of α over F, and let $n = \deg(m(x))$. We then have that $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is a basis for $F(\alpha)$ over F. Hence, $[F(\alpha) : F]$ is the degree of the minimal polynomial of α over F.*

Proof. Given any $h(x) \in F[x]$ with $\deg(h(x)) < n$, write $h(x) = c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ with $c_i \in F$ (where possibly some c_i are 0), and notice that

$$h(\alpha) = c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0$$

so $h(\alpha)$ is in the span of the set $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$. Since we have just shown that

$$F(\alpha) = F[\alpha] = \{0\} \cup \{h(\alpha) : h(x) \in F[x] \text{ and } \deg(h(x)) < n\}$$

it follows that $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ does indeed span $F(\alpha)$ over F. We now need only show that the set $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is linearly independent over F. Suppose that $c_i \in F$ with $c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 = 0$. Letting $h(x) = c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$, we then have that $h(\alpha) = 0$, so $h(x) \in \langle m(x) \rangle$. Since either $h(x) = 0$ or $\deg(h(x)) < n$ and the latter is impossible because $\deg(m(x)) = n$, we conclude that $h(x) = 0$. Therefore, $c_i = 0$ for all i. It follows that the set $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is linearly independent and hence a basis for $F(\alpha)$ over F. □

Proposition 4.2.10. *Suppose that $F \subseteq K$ is a field extension and that $\alpha \in K$ is transcendental over F.*

1. *The set $\{1, \alpha, \alpha^2, \alpha^3, \ldots\}$ is linearly independent over F.*
2. *$[F(\alpha) : F] = \infty$.*
3. *$F[x] \cong F[\alpha]$.*
4. *$F[\alpha] \subsetneq F(\alpha)$.*

Proof. Suppose that $n \in \mathbb{N}$ and $a_i \in F$ are such that

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n = 0$$

We then have that α is a root of the polynomial

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

Since α is transcendental over F, we must have that $p(x)$ is the zero polynomial, hence we must have $a_i = 0$ for all i. Therefore, $\{1, \alpha, \alpha^2, \alpha^3, \ldots\}$ is linearly independent over F. Since we have found an infinite subset of $F[\alpha]$ that is linearly independent over F, we can conclude that $[F(\alpha) : F] = \infty$.

We now prove the third statement. Recall the map $Ev_\alpha \colon F[x] \to K$ given by $Ev_\alpha(p(x)) = p(\alpha)$ is such that $\ker(Ev_\alpha) = I_\alpha$ and $\text{range}(Ev_\alpha) = F[\alpha]$. Since α is transcendental over F, we have $\ker(Ev_\alpha) = I_\alpha = \{0\}$. Since Ev_α is a ring homomorphism with trivial kernel, we must have that Ev_α is injective. Using the fact that $\text{range}(Ev_\alpha) = F[\alpha]$, it follows that $F[x] \cong F[\alpha]$ via the map Ev_α.

The fourth statement now follows because $F[\alpha] \cong F[x]$, so $F[\alpha]$ is not a field because $F[x]$ is not a field. □

Corollary 4.2.11. *Let $F \subseteq K$ be a field extension and let $\alpha \in K$. We then have that α is algebraic over F if and only if $F[\alpha]$ is finitely generated over F (i.e. if and only if $F[\alpha]$ can be spanned by a finite set using coefficients from F).*

Proof. Suppose that α is algebraic over F. Let $m(x) \in F[x]$ be the minimal polynomial of α over F. Letting $n = \deg(m(x))$, we then know that $F(\alpha) = F[\alpha]$ is generated by $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ over F.

Suppose that α is transcendental over F. Although $F[\alpha]$ is not a field, we can still view it as a vector space over F (there is no notion of "inverse" of a vector in a vector space). If $F[\alpha]$ is finitely generated over F, then it would have a finite spanning set. From basic linear algebra, if X is a finite spanning set and $|Y| > |X|$, then Y must be linearly dependent over F. However, we know that $F[\alpha]$ has an infinite linearly independent set, namely $\{1, \alpha, \alpha^2, \alpha^3, \ldots\}$. □

4.3 Irreducible Polynomials

Proposition 4.3.1. *Let F be a field and let $p(x) \in F[x]$. An element $a \in F$ is a root of $p(x)$ if and only if $x - a$ divides $p(x)$ in $F[x]$*

Proof. Suppose first that $x - a$ divides $p(x)$ in $F[x]$. Fix a polynomial $q(x) \in F[x]$ with $p(x) = (x - a)q(x)$. We then have that

$$p(a) = (a - a)q(a) = 0 \cdot q(a) = 0$$

so a is a root of $p(x)$.

Suppose conversely that a is a root of $p(x)$. Since $F[x]$ is a Euclidean domain with Euclidean function equal to the degree map, we may fix $q(x), r(x) \in F[x]$ with

$$p(x) = q(x)(x - a) + r(x)$$

and either $r(x) = 0$ or $\deg(r(x)) < \deg(x - a)$. Now $\deg(x - a) = 1$, so we have have that $r(x) = c$ is a constant polynomial. We then have

$$0 = p(a) = q(a)(a - a) + r(a) = q(a) \cdot 0 + c = c$$

so $r(x)$ is the zero polynomial. It follows that $p(x) = q(x)(x - a)$, so $x - a$ divides $p(x)$ in $F[x]$. □

Proposition 4.3.2. *Let F be a field and let $f(x) \in F[x]$ be a nonzero polynomial with $\deg(f(x)) \geq 2$. If $f(x)$ has a root in F, then $f(x)$ is not irreducible in $F[x]$.*

Proof. If $f(x)$ has a root a, then $(x - a) \mid f(x)$. Fixing $g(x) \in F[x]$ with $f(x) = (x - a) \cdot g(x)$. We then have

$$\deg(f(x)) = \deg(x - a) + \deg(g(x)) = 1 + \deg(g(x))$$

so $\deg(g(x)) = \deg(f(x)) - 1 \geq 1$. Now the units of $F[x]$ are the nonzero constants, so we have factored $f(x)$ as the product of two nonunits, and hence $f(x)$ is not irreducible in $F[x]$. □

Unfortunately, the test for the existence of roots is not in general sufficient to guarantee that a polynomial is irreducible, but it is enough in the special case where the polynomial has degree either 2 or 3.

Proposition 4.3.3. *Let F be a field and let $f(x) \in F[x]$ be a polynomial with either $\deg(f(x)) = 2$ or $\deg(f(x)) = 3$. If $f(x)$ has no roots in $F[x]$, then $f(x)$ is irreducible in $F[x]$.*

Proof. We prove the contrapositive. Suppose conversely that $f(x) \in F[x]$ is not irreducible. Write $f(x) = g(x)h(x)$ where $g(x), h(x) \in F[x]$ are nonunits. We have

$$\deg(f(x)) = \deg(g(x)) + \deg(h(x))$$

Now $g(x)$ and $h(x)$ are not units, so they each have degree at least 1. Since $\deg(f(x)) \in \{2, 3\}$, it follows that at least one of $g(x)$ or $h(x)$ has degree equal to 1. Suppose without loss of generality that $\deg(g(x)) = 1$ and write $g(x) = ax + b$ where $a, b \in F$ with $a \neq 0$. We then have

$$\begin{aligned} f(-ba^{-1}) &= g(-ba^{-1}) \cdot h(-ba^{-1}) \\ &= (a \cdot (-ba^{-1}) + b) \cdot h(-ba^{-1}) \\ &= (-b + b) \cdot h(-ba^{-1}) \\ &= 0 \cdot h(-ba^{-1}) \\ &= 0 \end{aligned}$$

so $-ba^{-1}$ is a root of $f(x)$. □

For example, consider the polynomial $f(x) = x^3 - 2$ over $\mathbb{Q}$. We know that $f(x)$ has no roots in $\mathbb{Q}$ because $\pm\sqrt[3]{2}$ are not rational (use the Fundamental Theorem of Arithmetic to prove this if you have have not seen it). Thus, $f(x)$ is irreducible over $\mathbb{Q}$. Notice that $f(x)$ is not irreducible when viewed as an element of $\mathbb{R}[x]$ because it has a root in $\mathbb{R}$. Moreover, no polynomial in $\mathbb{R}[x]$ of odd degree is irreducible because every such polynomial has a root (this uses the Intermediate Value Theorem because as $x \to \pm\infty$, on one side you must have $f(x) \to \infty$ and on the other you must have $f(x) \to -\infty$). In fact, it turns out that every irreducible polynomial over $\mathbb{R}$ has degree either 1 or 2, though this is far from obvious at this point since there are certainly polynomials of degree 4 with no root, such as $x^4 + 1$.

Proposition 4.3.4. *Let R be a ring and let I be an ideal of R. Define a function $\psi\colon R[x] \to (R/I)[x]$ by letting*

$$\psi(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0) = \overline{a_n}x^n + \overline{a_{n-1}}x^{n-1} + \cdots + \overline{a_1}x + \overline{a_0}$$

The function ψ is a surjective ring homomorphism with kernel $I[x]$, hence

$$R[x]/I[x] \cong (R/I)[x]$$

Proof. This is all just direct computation. □

Lemma 4.3.5. *Suppose that $g(x), h(x) \in \mathbb{Z}[x]$ and that $p \in \mathbb{Z}$ is a prime which divides all coefficients of $g(x) \cdot h(x)$. We then have that either p divides all coefficients of $g(x)$, or p divides all coefficients of $h(x)$.*

Proof 1 - Elementary Proof. Let $g(x) = \sum_i b_i x^i$, let $h(x) = \sum_i c_i x^i$, and let $g(x)h(x) = \sum_i a_i x^i$. We are supposing that $p \mid a_i$ for all i. Suppose the $p \nmid b_i$ for some i and also that $p \nmid c_i$ for some i (possibly different). Let k be least such that $p \nmid b_k$, and let ℓ be least such that $p \nmid c_\ell$. Notice that

$$a_{k+\ell} = \sum_{i=0}^{k+\ell} b_i c_{k+\ell-i} = b_k c_\ell + (\sum_{i=0}^{k-1} b_i c_{k+\ell-i}) + (\sum_{i=k+1}^{k+\ell} b_i c_{k+\ell-i})$$

hence

$$b_k c_\ell = a_{k+\ell} - (\sum_{i=0}^{k-1} b_i c_{k+\ell-i}) - (\sum_{i=k+1}^{k+\ell} b_i c_{k+\ell-i})$$

Now if $0 \leq i \leq k-1$, then $p \mid b_i$ by choice of k, hence $p \mid b_i c_{k+\ell-i}$. Also, if $k+1 \leq i \leq k+\ell$, then $k+\ell-i < \ell$, so $p \mid c_{k+\ell-i}$ by choice of ℓ, hence $p \mid b_i c_{k+\ell-i}$. Since $p \mid a_{k+\ell}$ by assumption, it follows that p divides every summand on the right hand side. Therefore, p divides the right hand side, and thus $p \mid b_k c_\ell$. Since p is prime, it follows that either $p \mid b_k$ or $p \mid c_\ell$, but both of these are impossible by choice of k and ℓ. Therefore, it must be the case that either $p \mid b_n$ for all n, or $p \mid c_n$ for all n. □

Proof 2 - Algebraic Proof. Passing to the quotient in $\mathbb{Z}/p\mathbb{Z}$, we have that $\overline{g(x) \cdot h(x)} = \overline{0}$ in $(\mathbb{Z}/p\mathbb{Z})[x]$, hence

$$\overline{g(x)} \cdot \overline{h(x)} = \overline{0}$$

Since $\mathbb{Z}/p\mathbb{Z}$ is an integral domain, we know that the $(\mathbb{Z}/p\mathbb{Z})[x]$ is an integral domain, and hence either $\overline{g(x)} = 0$ or $\overline{h(x)} = 0$. In the former case, every coefficient of $g(x)$ is divisible by p, while in the latter every coefficient of $h(x)$ is divisible by p. □

Proposition 4.3.6 (Gauss' Lemma)**.** *Suppose that $f(x) \in \mathbb{Z}[x]$ and that $g(x), h(x) \in \mathbb{Q}[x]$ with $f(x) = g(x)h(x)$. There exist polynomials $g^*(x), h^*(x) \in \mathbb{Z}[x]$ such that $f(x) = g^*(x)h^*(x)$ and both $\deg(g^*(x)) = \deg(g(x))$ and $\deg(h^*(x)) = \deg(h(x))$. In fact, there exist nonzero $s, t \in \mathbb{Q}$ with*

- $f(x) = g^*(x)h^*(x)$

- $g^*(x) = s \cdot g(x)$
- $h^*(x) = t \cdot h(x)$

Proof. If each of the coefficients of $g(x)$ and $h(x)$ happen to be integers, then we are happy. Suppose not. Let $a \in \mathbb{Z}$ be the least common multiple of the denominators of the coefficients of g, and let $b \in \mathbb{Z}$ be the least common multiple of the denominators of the coefficients of h. Let $d = ab$. Multiply both sides of $f(x) = g(x)h(x)$ through by d to "clear denominators" gives

$$d \cdot f(x) = (a \cdot g(x)) \cdot (b \cdot h(x))$$

where each of the three factors $d \cdot f(x)$, $a \cdot g(x)$, and $b \cdot h(x)$ is a polynomial in $\mathbb{Z}[x]$. We have at least one of $a > 1$ or $b > 1$, hence $d = ab > 1$.

Fix a prime divisor p of d. We then have that p divides all coefficients of $d \cdot f(x)$, so by the previous lemma either p divides all coefficients of $a \cdot g(x)$, or p divides all coefficients of $b \cdot h(x)$. In the former case, we have

$$\frac{d}{p} \cdot f(x) = (\frac{a}{p} \cdot g(x)) \cdot (b \cdot h(x))$$

where each of the three factors is a polynomial in $\mathbb{Z}[x]$. In the latter case, we have

$$\frac{d}{p} \cdot f(x) = (a \cdot g(x)) \cdot (\frac{b}{p} \cdot h(x))$$

where each of the three factors is a polynomial in $\mathbb{Z}[x]$. Now if $\frac{d}{p} = 1$, then we are done by letting $g^*(x)$ be the first factor and letting $h^*(x)$ be the second. Otherwise, we continue the argument by dividing out another prime factor of $\frac{d}{p}$ from all coefficients of one of the two polynomials. Continue until we have handled all primes which occur in a factorization of d. Formally, you can do induction on d. □

An immediate consequence of Gauss' Lemma is the following, which greatly simplifies the check for whether a given polynomial with integer coefficients is irreducible in $\mathbb{Q}[x]$.

Corollary 4.3.7. *Let $f(x) \in \mathbb{Z}[x]$. If there does not exist nonconstant polynomials $g(x), h(x) \in \mathbb{Z}[x]$ with $f(x) = g(x) \cdot h(x)$, then $f(x)$ is irreducible in $\mathbb{Q}[x]$. Furthermore, if $f(x)$ is monic, then it suffices to show that no such monic $g(x)$ and $h(x)$ exist.*

Proof. The first part is immediate from Gauss' Lemma. Now suppose that $f(x) \in \mathbb{Z}[x]$ is monic. Suppose that $g(x), h(x) \in \mathbb{Z}$ with $f(x) = g(x)h(x)$. Notice that the leading term of $f(x)$ is the product of the leading terms of $g(x)$ and $h(x)$, so as $f(x)$ is monic and all coefficients are in $\mathbb{Z}$, either both $g(x)$ and $h(x)$ are monic or both have leading terms -1. In the latter case, we can multiply both through by -1 to get a factorization into monic polynomials in $\mathbb{Z}[x]$ of the same degree. □

Proposition 4.3.8. *Suppose that $f(x) \in \mathbb{Z}[x]$ is monic and $p \in \mathbb{Z}$ is prime. Suppose that $\overline{f(x)}$ is an irreducible polynomial in $(\mathbb{Z}/p\mathbb{Z})[x]$. We then have that $f(x)$ is irreducible in $\mathbb{Q}[x]$.*

Proof. We prove the contrapositive. Suppose that $f(x)$ is reducible in $\mathbb{Q}[x]$. By Gauss' Lemma, there exist nonconstant polynomials $g(x), h(x) \in \mathbb{Z}[x]$ with $f(x) = g(x)h(x)$. Passing to the quotient and using the above isomorphism, it follows that

$$\overline{f(x)} = \overline{g(x)} \cdot \overline{h(x)}$$

in the ring $(\mathbb{Z}/p\mathbb{Z})[x]$. Now $\deg(\overline{f(x)}) = \deg(f(x))$ because $f(x)$ is monic. Since $\deg(\overline{g(x)}) \leq \deg(g(x))$ and $\deg(\overline{h(x)}) \leq \deg(h(x))$, we have shown that $\overline{f(x)}$ can be factored into two polynomials of smaller degree in $(\mathbb{Z}/p\mathbb{Z})[x]$, so $\overline{f(x)}$ is reducible there. □

Theorem 4.3.9 (Eisenstein's Criterion)**.** *Suppose that $f(x) \in \mathbb{Z}[x]$ is monic and write*

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

If there exists a prime $p \in \mathbb{Z}$ such that

- *$p \mid a_i$ for $0 \leq i \leq n-1$*
- *$p^2 \nmid a_0$*

then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Proof 1 - Elementary Proof. Fix such a prime p. We use Corollary 4.3.7. Suppose that $g(x), h(x) \in \mathbb{Z}[x]$ are not constant polynomials with $f(x) = g(x)h(x)$. We then have

$$n = \deg(f(x)) = \deg(g(x)) + \deg(h(x))$$

Since we are assuming that $g(x)$ and $h(x)$ are not constant, they each have degree at least 1, and so by the above equality they both have degree at most $n-1$.

Let $g(x) = \sum_i b_i x^i$, let $h(x) = \sum_i c_i x^i$. We have $a_0 = b_0 c_0$, so since $p \mid a_0$ and p is prime, either $p \mid b_0$ or $p \mid c_0$. Furthermore, since $p^2 \nmid a_0$ by assumption, we can not have both $p \mid b_0$ and $p \mid c_0$. Without loss of generality (by switching the roles of $g(x)$ and $h(x)$ if necessary), suppose that $p \mid b_0$ and $p \nmid c_0$.

We now prove that $p \mid b_k$ for $0 \leq k \leq n-1$ by (strong) induction. Suppose that we have k with $0 \leq k \leq n-1$ and we know that $p \mid b_i$ for $0 \leq i < k$. Now

$$a_k = b_k c_0 + b_{k-1}c_1 + \cdots + b_1 c_{k-1} + b_0 c_k$$

and hence

$$b_k c_0 = a_k - b_{k-1}c_1 - \cdots - b_1 c_{k-1} - b_0 c_k$$

By assumption, we have $p \mid a_k$, and by induction we have $p \mid b_i$ for $0 \leq i < k$. It follows that p divides every term on the right-hand side, so $p \mid b_k c_0$. Since p is prime and $p \nmid c_0$, it follows that $p \mid b_k$.

Thus, we have shown that $p \mid b_k$ for $0 \leq k \leq n-1$. Now we have

$$\begin{aligned} a_n &= b_n c_0 + b_{n-1}c_1 + \cdots + b_1 c_{n-1} + b_0 c_n \\ &= b_{n-1}c_1 + \cdots + b_1 c_{n-1} + b_0 c_n \end{aligned}$$

where the last line follows from the fact that $b_n = 0$ (since we are assuming $\deg(g(x)) < n$). Now we know $p \mid b_k$ for $0 \leq k \leq n-1$, so p divides every term on the right. It follows that $p \mid a_n$, contradicting our assumption. Therefore, by Corollary 4.3.7, $f(x)$ is irreducible in $\mathbb{Q}[x]$. □

Proof 2 - Algebraic Proof. Fix such a prime p. We use Corollary 4.3.7. Suppose that $g(x), h(x) \in \mathbb{Z}[x]$ are nonconstant with $f(x) = g(x)h(x)$. Passing to the quotient $\mathbb{Z}/p\mathbb{Z}$ and using the above isomorphism, it follows that

$$\overline{f(x)} = \overline{g(x)} \cdot \overline{h(x)}$$

in the ring $(\mathbb{Z}/p\mathbb{Z})[x]$. Now we have $p \mid a_i$ for $0 \leq i \leq n-1$, so $\overline{f(x)} = x^n$ in $(\mathbb{Z}/p\mathbb{Z})[x]$. Thus, in the ring $(\mathbb{Z}/p\mathbb{Z})[x]$, we have

$$x^n = \overline{g(x)} \cdot \overline{h(x)}$$

Now each of $g(x), h(x) \in \mathbb{Z}[x]$ have degree at most $n-1$, so this is certainly true of $\overline{g(x)}, \overline{h(x)} \in (\mathbb{Z}/p\mathbb{Z})[x]$ as well. We know that $(\mathbb{Z}/p\mathbb{Z})[x]$ is a Euclidean domain, hence a UFD, so as the polynomial x is irreducible in $(\mathbb{Z}/p\mathbb{Z})[x]$, it follows that the only divisors of x^n in $(\mathbb{Z}/p\mathbb{Z})[x]$ are the polynomials ux^k where u is a nonzero constant (unit) and $0 \leq k \leq n$. Thus, each of $\overline{g(x)}$ and $\overline{h(x)}$ are of this form, and since neither can be constant polynomials (since their product has degree n and each has degree at most $n-1$), it follows that each of $\overline{g(x)}$ and $\overline{h(x)}$ have constant term 0. Thus, the constant terms of $g(x)$ and $h(x)$ are each divisible by p. It follows that the constant term of $f(x)$ is divisible by p^2 which contradicts our assumption. Therefore, $f(x)$ has no nontrivial factorization in $\mathbb{Z}[x]$, and hence none in $\mathbb{Q}[x]$ either by Gauss' Lemma. □

Example 4.3.10. *Let $m \geq 2$ be a squarefree integer (so no nontrivial square divides m, i.e. all primes in the factorization of m are distinct). Let $n \geq 2$. The polynomial $x^n - m$ is irreducible in $\mathbb{Q}[x]$ because it satisfies Eisenstein's Criterion with p chosen to be any prime which divides m. Thus, $x^n - m$ is the minimal polynomial of $\sqrt[n]{m}$ over $\mathbb{Q}$.*

Corollary 4.3.11. *If $m \geq 2$ is a squarefree integer and $n \geq 2$, then $[\mathbb{Q}(\sqrt[n]{m}) : \mathbb{Q}] = n$.*

4.4 Finite and Algebraic Extensions

Definition 4.4.1. *Let $F \subseteq K$ be a field extension.*

- *We say that the extension $F \subseteq K$ is* finite *if $[K : F] < \infty$.*
- *We say that the extension $F \subseteq K$ is* algebraic *if every $\alpha \in K$ is algebraic over F.*

Theorem 4.4.2. *Every finite extension $F \subseteq K$ is algebraic.*

Proof. Suppose that $[K : F]$ is finite and let $n = [K : F]$. Let $\alpha \in K$. Now the set $\{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ is a set of $n + 1$ vectors in the vector space K over F, so as $\dim_F K = n$, it follows that this set must be linearly dependent over F. Therefore, there exists $c_i \in F$ not all zero such that

$$c_n\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 = 0$$

Letting $p(x) = c_nx^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0 \in F[x]$, we have that $p(x)$ is a nonzero polynomial and $p(\alpha) = 0$, so α is algebraic over F. Since $\alpha \in K$ was arbitrary, the result follows. □

Notice that if $F \subseteq K$ and $\alpha \in K$ is algebraic over F, then the field extension $[F(\alpha) : F]$ is finite. Thus, every element of $F(\alpha)$ is algebraic over F.

Definition 4.4.3. *Let $F \subseteq K$ be a field extension and let $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$. The set $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is defined to be the smallest subfield of K containing $F \cup \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$.*

It is straightforward to check that

$$F(\alpha, \beta) = (F(\alpha))(\beta)$$

so we can get $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ by repeatedly adjoining one element.

Theorem 4.4.4. *Suppose that $F \subseteq K \subseteq L$ are field extensions. If both $[L : K]$ and $[K : F]$ are finite, then so is $[L : F]$ and we have*

$$[L : F] = [L : K] \cdot [K : F]$$

If either $[L : K]$ or $[K : F]$ is infinite, then so is $[L : F]$.

Proof. If $[L : K]$ is infinite, then there exists an infinite subset of L which is linearly independent over K, and this set is then trivially linearly independent over F, so $[L : F]$ is infinite. If $[K : F]$ is infinite, then there exists an infinite subset of K which is linearly independent over F, and since $K \subseteq L$, this is an infinite subset of L which is linearly independent over F, so $[L : F]$ is infinite.

Suppose then that $m = [L : K]$ and $n = [K : F]$ are both finite. Let $\{\beta_1, \beta_2, \ldots, \beta_m\}$ be a basis of L over K, and let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a basis of K over F. We claim that the set

$$Z = \{\alpha_i\beta_j : 1 \leq i \leq n, 1 \leq j \leq m\}$$

is a basis for L over F. Let $\gamma \in L$. Since $\{\beta_1, \beta_2, \ldots, \beta_m\}$ is a basis of L over K, there exists $d_1, d_2, \ldots, d_m \in K$ with

$$\gamma = d_1\beta_1 + d_2\beta_2 + \cdots + d_m\beta_m$$

Now each $d_j \in K$, so as $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a basis of K over F, there exists $c_{1j}, c_{2j}, \ldots, c_{nj} \in F$ with

$$d_j = c_{1j}\alpha_1 + c_{2j}\alpha_2 + \cdots + c_{nj}\alpha_n$$

We then have

$$\begin{aligned}\gamma &= \sum_{j=1}^{m} d_j\beta_j \\ &= \sum_{j=1}^{m}(\sum_{i=1}^{n} c_{ij}\alpha_i)\beta_j \\ &= \sum_{j=1}^{m}\sum_{i=1}^{n} c_{ij}\alpha_i\beta_j\end{aligned}$$

Since $\gamma \in L$ was arbitrary, it follows that the set Z spans L over F.

We now need to check that Z is linearly independent over F. Suppose that $c_{ij} \in F$ satisfy

$$\sum_{j=1}^{m}\sum_{i=1}^{n} c_{ij}\alpha_i\beta_j = 0$$

We then have

$$\sum_{j=1}^{m}(\sum_{i=1}^{n} c_{ij}\alpha_i)\beta_j$$

so as $\sum_{i=1}^{n} c_{ij}\alpha_i \in K$ and $\{\beta_1, \beta_2, \ldots, \beta_m\}$ is linearly independent over K, we can conclude that $\sum_{i=1}^{n} c_{ij}\alpha_i = 0$ for each j. Now each $c_{ij} \in F$ and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is linearly independent over F, so we can conclude that each $c_{ij} = 0$. It follows that Z is linearly independent over F.

Combining the above, we have shown that $Z \subseteq L$ spans L over F and is linearly independent over F, so Z is a basis of L over F. Therefore,

$$[L : F] = |Z| = mn = [L : K] \cdot [K : F]$$

□

Corollary 4.4.5. *Let $F \subseteq K$. If $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$ are all algebraic over F, then $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a finite (and hence algebraic) extension of F.*

Proof. We have

$$F \subseteq F(\alpha_1) \subseteq F(\alpha_1, \alpha_2) \subseteq \cdots \subseteq F(\alpha_1, \alpha_2, \ldots, \alpha_n)$$

Each particular extension is finite, so the tower is finite. □

Proposition 4.4.6. *Let $F \subseteq K$ be a field extension. Let*

$$A = \{\alpha \in K : \alpha \text{ is algebraic over } F\}$$

We then have that A is subfield of K. Furthermore, A is an algebraic extension of F.

Proof. Suppose that $\alpha, \beta \in A$. We then have that α and β are both algebraic over F, so $[F(\alpha, \beta) : F]$ is finite. Therefore, every element of $F(\alpha, \beta)$ is algebraic over F. Since $\alpha + \beta$, $\alpha - \beta$, $\alpha \cdot \beta$, and $\frac{\alpha}{\beta}$ (if $\beta \neq 0$) are all elements of $F(\alpha, \beta)$, it follows that each of these are algebraic over F. Thus, $\alpha + \beta$, $\alpha - \beta$, $\alpha \cdot \beta$, and $\frac{\alpha}{\beta}$ (if $\beta \neq 0$) are all in A. □

As an example, consider $L = \mathbb{Q}(\sqrt{2}, i)$. Notice that if $K = \mathbb{Q}(\sqrt{2})$, then we have $\mathbb{Q} \prec K \prec L$. Now $[K : \mathbb{Q}] = 2$ and $K \neq L$, so $[L : K] \neq 1$. It follows that $[L : K] = 2$, and thus $[L : \mathbb{Q}] = 4$. A basis is $\{1, \sqrt{2}, i, \sqrt{2}i\}$. Now we claim that $L = \mathbb{Q}(\sqrt{2} + i)$. It can be shown that $\sqrt{2} + i$ is a root of the polynomial $x^4 - 2x^2 + 9$. It is possible to show that is irreducible directly, but it is painful. If you do this, then you know that $[\mathbb{Q}(\sqrt{2} + i) : \mathbb{Q}] = 4$, so $[\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}(\sqrt{2} + i)] = 1$, so they are equal.

We can also argue from the other direction. Notice that

$$\begin{aligned}(\sqrt{2} + i)^3 &= (\sqrt{2})^3 + 3 \cdot (\sqrt{2})^2 \cdot i + 3 \cdot \sqrt{2} \cdot i^2 + i^3 \\ &= 2\sqrt{2} + 6i - 3\sqrt{2} - i \\ &= -\sqrt{2} + 5i\end{aligned}$$

It follows that

$$(\sqrt{2} + i)^3 + (\sqrt{2} + i) = 6i$$

and therefore

$$i = \frac{1}{6}(\sqrt{2} + i)^3 + \frac{1}{6}(\sqrt{2} + i)$$

Thus, $i \in \mathbb{Q}(\sqrt{2} + i)$ and it follows that $\sqrt{2} \in \mathbb{Q}(\sqrt{2} + i)$ as well. Hence, $\mathbb{Q}(\sqrt{2} + i) = \mathbb{Q}(\sqrt{2}, i)$, and so $[\mathbb{Q}(\sqrt{2} + i) : \mathbb{Q}] = 4$. Therefore, the fourth degree polynomial $x^4 - 2x^2 + 9$ must be the minimal polynomial of $\sqrt{2} + i$ over $\mathbb{Q}$.

Proposition 4.4.7. *Suppose that $F \subseteq E \subseteq K$ and that E is algebraic over F. If $\alpha \in K$ and α is algebraic over E, then α is algebraic over F.*

Proof. Suppose $\alpha \in K$ and α is algebraic over E. Fix a polynomial $p(x) \in E[x]$ with $p(\alpha) = 0$. Write

$$p(x) = \beta_n x^n + \beta_{n-1} x^{n-1} + \cdots + \beta_1 x + \beta_0 \in E[x]$$

where each $\beta_i \in E$. Since E is algebraic over F, we know that each β_i is algebraic over F. Therefore, the field $L = F(\beta_1, \beta_2, \ldots, \beta_n)$ is such that $[L : F]$ is finite. Notice that $p(x) \in L[x]$, so α is algebraic over L and thus $[L(\alpha) : L]$ is finite. It follows that $[L(\alpha) : F]$ is finite, and therefore $\alpha \in L(\alpha)$ is algebraic over F. □

For example, any root of the polynomial

$$(\sqrt[4]{5})x^9 - (\sqrt{2} + 6i)x^5 + (e^{2\pi i/5} \cdot \sqrt[3]{2})$$

must be algebraic over $\mathbb{Q}$

Corollary 4.4.8. *Suppose that $F \subseteq E \subseteq K$. If E is algebraic over F and K is algebraic over E, then K is algebraic over F.*

4.5 Algebraic Integers

Definition 4.5.1. *A number field is a field K such that $\mathbb{Q} \subseteq K \subseteq \mathbb{C}$ with the property that $[K : \mathbb{Q}]$ is finite.*

Suppose that we have a number field K. As $[K : \mathbb{Q}]$ is finite, we know that every element of K is algebraic over $\mathbb{Q}$. Since we have extended the field $\mathbb{Q}$ to a slightly larger field K, we might think that the analogue of the integers in K will be a slightly larger ring than $\mathbb{Z}$. The fundamental question is how to define this "ring of integers" in K. If there exists $\alpha \in \mathbb{C}$ with $K = \mathbb{Q}(\alpha)$, it is very natural to expect that the proper analogue would be

$$\mathbb{Z}[\alpha] = \{p(\alpha) : p(x) \in \mathbb{Z}[x]\}$$

After all, this is easily seen to be the smallest subring of $\mathbb{Q}(\alpha)$ containing $\mathbb{Z} \cup \{\alpha\}$. Furthermore, it matches our expectation that the Gaussian Integers $\mathbb{Z}[i]$ are the correct analogue of the "integers" in the number field $\mathbb{Q}(i)$. However, it is problematic for a few reasons, as we now explore.

Lemma 4.5.2. *Suppose that $F \subseteq K$ is a field extension and let $\alpha, \beta \in K$.*

- *$\alpha \in F(\beta)$ if and only if $F(\alpha) \subseteq F(\beta)$.*
- *We have both $\alpha \in F(\beta)$ and $\beta \in F(\alpha)$ if and only if $F(\alpha) = F(\beta)$.*

Proof. If $F(\alpha) \subseteq F(\beta)$, then since $\alpha \in F(\alpha)$ we clearly have $\alpha \in F(\beta)$. Conversely, if $\alpha \in F(\beta)$, then $F(\beta)$ is a subfield of K containing $F \cup \{\alpha\}$, so as $F(\alpha)$ is the smallest such subfield it follows that $F(\alpha) \subseteq F(\beta)$. The second statement is immediate from the first. □

Consider the field $\mathbb{Q}(\sqrt{-3})$ and $\mathbb{Q}(\zeta)$ where $\zeta = e^{2\pi i/3}$. Notice that $\sqrt{-3}$ is algebraic over $\mathbb{Q}$ with minimal polynomial x^2+3 (this is irreducible over $\mathbb{Q}$ because it has no roots in $\mathbb{Q}$). Now ζ is a root of the polynomial x^3-1. However, this polynomial is not irreducible over $\mathbb{Q}$ because

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

Since ζ is a root of x^3-1, it must be a root of one of the polynomials on the right, and since $\zeta \neq 1$, it follows that ζ is a root of x^2+x+1. One can check that this polynomial has no rational roots (or even real root), so it is irreducible in $\mathbb{Q}[x]$ and hence the minimal polynomial of ζ over $\mathbb{Q}$. Now using Euler's Formula

$$e^{i\theta} = \cos\theta + i\sin\theta$$

we see that

$$\begin{aligned}\zeta &= e^{2\pi i/3} \\ &= \cos(2\pi/3) + i\sin(2\pi/3) \\ &= -\frac{1}{2} + i\frac{\sqrt{3}}{2} \\ &= -\frac{1}{2} + \frac{\sqrt{-3}}{2}\end{aligned}$$

With this in hand, we claim that $\mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta)$. To see this, notice that the above equality shows that $\zeta \in \mathbb{Q}(\sqrt{-3})$ and hence $\mathbb{Q}(\zeta) \subseteq \mathbb{Q}(\sqrt{-3})$. Since

$$\sqrt{-3} = 2\zeta + 1$$

we also have $\sqrt{-3} \in \mathbb{Q}(\zeta)$. Therefore $\mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta)$.

Definition 4.5.3. *Given $\alpha \in \mathbb{C}$, we let $\mathbb{Z}[\alpha]$ be the smallest subring of $\mathbb{C}$ containing α (notice that any subring contains 1, and hence all of $\mathbb{Z}$). In other words,*

$$\mathbb{Z}[\alpha] = \{p(\alpha) : p(x) \in \mathbb{Z}[x]\}$$

Since $\mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta)$, should the "integers" of this field be

$$\mathbb{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbb{Z}\}$$

or the larger ring

$$\begin{aligned}
\mathbb{Z}[\zeta] &= \{a + b\zeta : a, b \in \mathbb{Z}\} \\
&= \{a + b(-\frac{1}{2} + \frac{\sqrt{-3}}{2}) : a, b \in \mathbb{Z}\} \\
&= \{(a - \frac{b}{2}) + \frac{b}{2}\sqrt{-3} : a, b \in \mathbb{Z}\} \\
&= \{\frac{2a - b}{2} + \frac{b}{2}\sqrt{-3} : a, b \in \mathbb{Z}\} \\
&= \{c + d\sqrt{-3} : c, d \in \mathbb{Z}\} \cup \{\frac{c}{2} + \frac{d}{2}\sqrt{-3} : c, d \in \mathbb{Z} \text{ both odd}\} \\
&= \mathbb{Z}[\sqrt{-3}] \cup \{\frac{c}{2} + \frac{d}{2}\sqrt{-3} : c, d \in \mathbb{Z} \text{ both odd}\}
\end{aligned}$$

or something else entirely?

The answer to this question is not immediately obvious. Notice that every element of $K = \mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta)$ is algebraic over $\mathbb{Q}$, so is the root of some nonzero polynomial in $\mathbb{Q}[x]$. A first guess for the definition of an "algebraic integer" (rather than just an algebraic number) would be that it is a root of nonzero polynomial in $\mathbb{Z}[x]$. However, by multiplying a polynomial in $\mathbb{Q}[x]$ through by the product of the denominators of coefficients, it is easy to see that every algebraic $\alpha \in \mathbb{C}$ is a root of a nonzero polynomial in $\mathbb{Z}[x]$. Thus, simply forcing the coefficients to be elements of $\mathbb{Z}$ does not change anything.

Definition 4.5.4. *Let $\alpha \in \mathbb{C}$. We say that α is an* algebraic integer *if α is the root of some nonzero* monic *polynomial in $\mathbb{Z}[x]$.*

Example 4.5.5. *The following are examples of algebraic integers*

- *Every $n \in \mathbb{Z}$ is an algebraic integer because it is a root of $x - n$.*
- *$\sqrt{2}$ is an algebraic integer because it is a root of $x^2 - 2$.*
- *i is an algebraic integer because it is a root of $x^2 + 1$.*
- *For every $n \geq 2$, the complex number $e^{\frac{2\pi i}{n}}$ is an algebraic integer because it is a root of $x^n - 1$.*
- *$\sqrt{2} + \sqrt{3}$ is an algebraic integer because it is a root of $x^4 - 10x^2 + 1$.*

Why is this the "correct" definition. There is no short answer to this question, but we begin with another characterization of algebraic integers that provides the first argument for it being the right choice. Recall that $\alpha \in \mathbb{C}$ is algebraic over $\mathbb{Q}$ if and only if

$$\mathbb{Q}[\alpha] = \{p(\alpha) : p(x) \in \mathbb{Q}\}$$

is finitely generated over $\mathbb{Q}$ (i.e. there is a finite set that spans $\mathbb{Q}[\alpha]$ over $\mathbb{Q}$).

Theorem 4.5.6. *Let $\alpha \in \mathbb{C}$. The following are equivalent.*

- *α is an algebraic integer.*
- *$\mathbb{Z}[\alpha]$ is a finitely generated additive subgroup of $\mathbb{C}$ (i.e. is finitely generated using only coefficients in $\mathbb{Z}$).*
- *There exists a subring R of $\mathbb{C}$ with $\alpha \in R$ such that R is a finitely generated additive group.*

Proof. $1 \to 2$: Suppose that α is an algebraic integer. Fix a monic polynomial $h(x) \in \mathbb{Z}[x]$ such that $h(\alpha) = 0$, say

$$h(x) = x^n + c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \cdots + c_1 x + c_0$$

One proof is that given any $p(x) \in \mathbb{Z}[x]$, there exists $q(x), r(x) \in \mathbb{Z}[x]$ with $p(x) = q(x)h(x) + r(x)$ and either $r(x) = 0$ or $\deg(r(x)) < \deg(h(x))$ (although $\mathbb{Z}$ isn't a field, the only thing you need in the proof for this to work in $F[x]$ is the fact that the leading coefficient is a unit in the ring). Thus, $p(\alpha) = r(\alpha)$, and we are done.

However, we now give a direct proof without using this fact. We show that $\alpha^m \in \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\rangle$ (viewed as the additive *subgroup* these generate) for all $m \geq n$ by induction. Since $h(\alpha) = 0$, it follows that

$$\alpha^n = -(c_0 \cdot 1 + c_1\alpha + \cdots + c_{n-2}\alpha^{n-2} + c_{n-1}\alpha^{n-1})$$

Thus, $\alpha^n \in \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\rangle$, so the base case is true. Suppose that $m \geq n$ and we have shown that $\alpha^m \in \langle 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\rangle$. Fix $k_0, k_1, \ldots, k_{n-1} \in \mathbb{Z}$ with

$$\alpha^m = k_0 \cdot 1 + k_1\alpha + \cdots + k_{n-2}\alpha^{n-2} + k_{n-1}\alpha^{n-1}$$

We then have

$$\begin{aligned}
\alpha^{m+1} &= \alpha^m \cdot \alpha \\
&= (k_0 \cdot 1 + k_1\alpha + \cdots + k_{n-2}\alpha^{n-2} + k_{n-1}\alpha^{n-1}) \cdot \alpha \\
&= k_0\alpha + k_1\alpha^2 + \cdots + k_{n-2}\alpha^{n-1} + k_{n-1}\alpha^n \\
&= (k_0\alpha + k_1\alpha^2 + \cdots + k_{n-2}\alpha^{n-1}) + k_{n-1} \cdot -(c_0 \cdot 1 + c_1\alpha + \cdots + c_{n-2}\alpha^{n-2} + c_{n-1}\alpha^{n-1}) \\
&= (-k_{n-1}c_0) \cdot 1 + (k_0 - k_{n-1}c_1) \cdot \alpha + \cdots + (k_{n-2} - k_{n-1}c_{n-1}) \cdot \alpha^{n-1}
\end{aligned}$$

Thus, the result holds for α^{m+1}. By induction, we see that α^m is an element of the additive subgroup generated by $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ for all m. From here, it follows that $\mathbb{Z}[\alpha]$ is generated as an additive abelian group by the set $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$.

$2 \to 3$: Trivial.

$3 \to 2$: Fix a subring R of $\mathbb{C}$ with $\alpha \in R$ such that R is a finitely generated additive group. Since R is a subring of $\mathbb{C}$, we have $1 \in R$ and hence $\mathbb{Z} \subseteq R$. Fix $\beta_1, \beta_2, \ldots, \beta_n$ which generate R as an additive abelian group, i.e. such that

$$R = \{k_1\beta_1 + k_2\beta_2 + \cdots + k_n\beta_n : k_i \in \mathbb{Z}\}$$

Since R is a ring, we know that $\alpha\beta_i \in R$ for each i, and hence there exists $k_{i,j} \in \mathbb{Z}$ such that

$$\alpha\beta_i = \sum_{j=1}^{n} k_{i,j}\beta_j$$

If we let M be the $n \times n$ matrix $M = [k_{i,j}]$ and let $\mathbf{v}$ be the $n \times 1$ column vector $\mathbf{v} = [\beta_i]$, then the above equation simply says that $\alpha\mathbf{v} = M\mathbf{v}$. Now $\mathbf{v} \neq \mathbf{0}$ because $R \neq \{0\}$ (we certainly have $1 \in R$), so α is an eigenvalue of M. Let $f(x) = \det(xI - M)$ be the characteristic polynomial of M. Notice that $f(x)$ is a monic polynomial (of degree n) and that $f(x) \in \mathbb{Z}[x]$ because all entries in M are integers. Since α is an eigenvalue of A, we know that α is a root of the characteristic polynomial $f(x)$, so α is an algebraic integer. □

Corollary 4.5.7. *If α and β are algebraic integers, then $\alpha + \beta$, $\alpha - \beta$, and $\alpha\beta$ are all algebraic integers. Therefore, the set of all algebraic integers is a subring of $\mathbb{C}$.*

Proof. Suppose that $\alpha, \beta \in \mathbb{C}$ are both algebraic integers. We claim that the ring

$$\mathbb{Z}[\alpha, \beta] = \{p(\alpha, \beta) : p(x, y) \in \mathbb{Z}[x, y]\}$$

is a finitely generated additive group. Fix monic polynomials $g(x), h(x) \in \mathbb{Z}[x]$ such that $g(\alpha) = 0$ and $h(\alpha) = 0$. Let $m = \deg(g(x))$ and let $n = \deg(h(x))$. As in the proof of the previous theorem, any α^k is in the additive subgroup generated by $\{1, \alpha, \alpha^2, \ldots, \alpha^{m-1}\}$ and any β^ℓ is in the additive subgroup generated by $\{1, \beta, \beta^2, \ldots, \beta^{n-1}\}$. It follows that any $\alpha^k \beta^\ell$ is an element of the additive subgroup generated by

$$\{\alpha^i \beta^j : 0 \leq i \leq m-1, 0 \leq j \leq n-1\}$$

It follows that this set generated $\mathbb{Z}[\alpha, \beta]$, so $\mathbb{Z}[\alpha, \beta]$ is finitely generated as an additive group.

Now each of $\alpha+\beta$, $\alpha-\beta$, and $\alpha\beta$ are elements of $\mathbb{Z}[\alpha, \beta]$, so they are all algebraic integers by the previous theorem. □

Definition 4.5.8. *Let K be a number field. We let $\mathcal{O}_K$ be the ring of all algebraic integers in K (the set is indeed a subring of K by the previous corollary).*

Theorem 4.5.9 (Rational Root Theorem)**.** *Suppose that $p(x) \in \mathbb{Z}[x]$ is a nonzero polynomial and write*

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where each $a_i \in \mathbb{Z}$ and $a_n \neq 0$. Suppose that $q \in \mathbb{Q}$ is a root of $p(x)$. If $q = \frac{b}{c}$ where $b, c \in \mathbb{Z}$ are relatively prime (so we write q in "lowest terms"), then $b \mid a_0$ and $c \mid a_n$.

Proof. We have

$$a_n \cdot (b/c)^n + a_{n-1} \cdot (b/c)^{n-1} + \cdots + a_1 \cdot (b/c) + a_0 = 0$$

Multiplying through by c^n we get

$$a_n b^n + a_{n-1} b^{n-1} c + \cdots + a_1 b c^{n-1} + a_0 c^n = 0$$

From this, we see that

$$a_n b^n = c \cdot [-(a_{n-1} b^{n-1} + \cdots + a_1 b c^{n-2} + a_0 c^{n-1})]$$

and hence $c \mid a_n b^n$. Using the fact that $\gcd(b, c) = 1$, it follows that $c \mid a_n$. On the other hand, we see that

$$a_0 c^n = b \cdot [-(a_n b^{n-1} + a_{n-1} b^{n-2} c + \cdots + a_1 c^{n-1})]$$

and hence $b \mid a_0 c^n$. Using the fact that $\gcd(b, c) = 1$, it follows that $b \mid a_0$. □

Corollary 4.5.10. *If $q \in \mathbb{Q}$ is an algebraic integer, then $q \in \mathbb{Z}$.*

Proof. Let $q \in \mathbb{Q}$ be an algebraic integer, and write $q = \frac{b}{c}$ where $b, c \in \mathbb{Z}$ and $\gcd(b, c) = 1$. Fix a nonzero monic polynomial $p(x) \in \mathbb{Z}[x]$ such that $p(q) = 0$. Since $p(x) \in \mathbb{Z}[x]$ is monic, we may write

$$p(x) = x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

Thus, by the previous Theorem, we have that $c \mid 1$ and $b \mid a_0$. It follows that $c \in \{-1, 1\}$, and hence $q = \frac{b}{c} = \pm b \in \mathbb{Z}$. □

Theorem 4.5.11. *Suppose that $\alpha \in \mathbb{C}$ is algebraic (over $\mathbb{Q}$). We then have that α is an algebraic integer if and only if the minimal polynomial of α over $\mathbb{Q}$ has integer coefficients.*

Proof. Since the minimal polynomial of α over $\mathbb{Q}$ is monic by definition, every such α is indeed an algebraic integer. Suppose conversely that α is an algebraic integer. Let $m(x)$ be the minimal polynomial of α over $\mathbb{Q}$. Since α is an algebraic integer, we may fix a monic nonzero $p(x) \in \mathbb{Z}[x]$ such that α is a root of $p(x)$. Since $m(x)$ is the minimal polynomial of α over $\mathbb{Q}$, it follows that $m(x)$ divides $p(x)$ in $\mathbb{Q}[x]$. Fix $h(x) \in \mathbb{Q}[x]$ with

$$p(x) = m(x) \cdot h(x)$$

Notice that by looking at leading terms and using the fact that both $p(x)$ and $m(x)$ are monic, we can conclude that $h(x)$ is also monic. By Gauss' Lemma, there exist $s, t \in \mathbb{Q}$ such that $s \cdot m(x) \in \mathbb{Z}[x]$, $t \cdot h(x) \in \mathbb{Z}[x]$ and

$$p(x) = (s \cdot m(x)) \cdot (t \cdot h(x))$$

Since $m(x)$ is monic and $s \cdot m(x) \in \mathbb{Z}[x]$, we must have that $s \in \mathbb{Z}$. Similarly, since $h(x)$ is monic and $t \cdot h(x) \in \mathbb{Z}[x]$, we must have that $t \in \mathbb{Z}$. Looking at leading terms, it follows that $st = 1$. Therefore, either $s = 1 = t$ or $s = -1 = t$. In either case, using the fact that $s \cdot m(x) \in \mathbb{Z}[x]$, we conclude that $m(x) \in \mathbb{Z}[x]$. $\square$

Chapter 5

Quadratic Number Fields

5.1 Classifying Quadratic Number Fields

Definition 5.1.1. *A* quadratic number field *is a number field K with $[K : \mathbb{Q}] = 2$.*

Definition 5.1.2. *An integer $d \in \mathbb{Z}$ is square-free if $ord_p(d) \leq 1$ for all primes p.*

The first few positive square-free numbers are $1, 2, 3, 5, 6, 7, 10, \ldots$ and the first few negative square-free numbers are $-1, -2, -3, -5, -6, -7, -10, \ldots$. Notice that d is square-free if and only if $-d$ is square-free.

Lemma 5.1.3. *For all $n \in \mathbb{Z}\backslash\{0\}$, there exist unique $m \in \mathbb{N}^+$ and a square-free $d \in \mathbb{Z}$ such that $n = m^2 d$.*

Proof. We first prove existence. If $n = 1$, we may take $m = 1$ and $d = 1$. If $n = -1$, we may take $m = 1$ and $d = -1$.

Let $P = \{p \in \mathbb{P} : ord_p(n)$ is even and nonzero$\}$ and let $Q = \{p \in \mathbb{P} : ord_p(n)$ is odd$\}$. Let

$$d = \prod_{p \in Q} p$$

and let

$$m = \left(\prod_{p \in P} p^{ord_p(n)/2} \right) \cdot \left(\prod_{p \in Q} p^{(ord_p(n)-1)/2} \right)$$

A simple check shows that $ord_p(m^2 d) = ord_p(n)$ for all primes p, so n and $m^2 d$ are associates. If $n = m^2 d$, we are done. If not, then $n = m^2(-d)$ and we are done since $-d$ is also square-free.

We now prove uniqueness. Let $n \in \mathbb{Z}\backslash\{0\}$. Suppose that $n = m^2 d$ and $n = \ell^2 c$ where $m, \ell \in \mathbb{N}^+$ and $c, d \in \mathbb{Z}$ are square-free. We have $\ell^2 c = m^2 d$. Since both $\ell^2 > 0$ and $m^2 > 0$, it follows that c and d are either both positive or both negative. Let $p \in \mathbb{N}^+$ be prime. Let $p \in \mathbb{N}^+$ be an arbitrary prime. We have

$$2 \cdot ord_p(\ell) + ord_p(c) = 2 \cdot ord_p(m) + ord_p(d)$$

so $ord_p(c) - ord_p(d)$ is even. Since c and d are both square-free, we know that $ord_p(c)$ and $ord_p(d)$ are both elements of $\{0, 1\}$. Thus, we must have $ord_p(c) = ord_p(d)$ and hence $ord_p(c) = ord_p(d)$. Since this is true for all primes $p \in \mathbb{N}^+$, it follows that c and d are associates in $\mathbb{Z}$. Combining this with the fact that c and d have the same sign, we conclude that $c = d$. Canceling this nonzero number from both sides of $\ell^2 c = m^2 d$, we deduce that $\ell^2 = m^2$. Since $\ell, m \in \mathbb{N}^+$, it follows that $\ell = m$. □

Theorem 5.1.4. *A number field K is a quadratic number field if and only if there exists a square-free $d \in \mathbb{Z}\backslash\{1\}$ with $K = \mathbb{Q}(\sqrt{d})$.*

Proof. Suppose first that $d \in \mathbb{Z}$ is square-free with $d \neq 1$. If $d = -1$, then $\sqrt{d} = i$ is a root of the irreducible $x^2 + 1 \in \mathbb{Q}[x]$, so $\mathbb{Q}(i)$ is a quadratic number field. Suppose then that $|d| > 2$. We have that $x^2 - d \in \mathbb{Q}[x]$ is irreducible in $\mathbb{Q}[x]$ by Eisenstein's Criterion applied to any prime divisor of d (or by directly checking that the roots are not rational), so $\mathbb{Q}(\sqrt{d})$ is a quadratic number field.

We prove the converse through a sequence of steps.

- We first show that there exists $r \in \mathbb{Q}$ with $K = \mathbb{Q}(\sqrt{r})$. Since $[K : \mathbb{Q}] > 1$, we know that $K \neq \mathbb{Q}$. Thus, we may fix $u \in K \backslash \mathbb{Q}$. Notice that $\mathbb{Q} \subsetneq \mathbb{Q}(u)$ because $\{1, u\}$ is linearly independent over $\mathbb{Q}$. Since $[K : \mathbb{Q}] = 2$ and $\mathbb{Q}(u)$ is a 2-dimensional subspace, we must have $K = \mathbb{Q}(u)$. Let $m(x) \in \mathbb{Q}[x]$ be the minimal polynomial of u over $\mathbb{Q}$ and write $m(x) = x^2 + bx + c$ where $b, c \in \mathbb{Q}$. Using the quadratic formula, we then have that either

$$u = \frac{-b + \sqrt{b^2 - 4c}}{2} \qquad \text{or} \qquad u = \frac{-b - \sqrt{b^2 - 4c}}{2}$$

  In either case, we have $u \in \mathbb{Q}(\sqrt{b^2 - 4c})$ because $b \in \mathbb{Q}$. In the former case we have $\sqrt{b^2 - 4c} = 2u + b \in \mathbb{Q}(u)$, while in the latter we have $\sqrt{b^2 - 4c} = -2u - b \in \mathbb{Q}(u)$ (again because $b \in \mathbb{Q}$), so in either case we have $\sqrt{b^2 - 4c} \in \mathbb{Q}(u)$. It follows that $\mathbb{Q}(u) = \mathbb{Q}(\sqrt{b^2 - 4c})$. Letting $r = b^2 - 4c \in \mathbb{Q}$, we are done.

- We now claim that there exists $n \in \mathbb{Z}$ with $K = \mathbb{Q}(\sqrt{n})$. Fix $r \in \mathbb{Q}$ with $K = \mathbb{Q}(\sqrt{r})$. Fix $a \in \mathbb{Z}$ and $b \in \mathbb{N}^+$ with $r = \frac{a}{b}$. We claim that $K = \mathbb{Q}(\sqrt{ab})$. Since $\sqrt{ab} = b \cdot \sqrt{\frac{a}{b}} = b\sqrt{r}$, we know that $\sqrt{ab} \in \mathbb{Q}(\sqrt{r})$. Since $\sqrt{r} = \sqrt{\frac{a}{b}} = \frac{1}{b} \cdot \sqrt{ab}$, we know that $\sqrt{r} \in \mathbb{Q}(\sqrt{ab})$. Therefore, letting $n = ab \in \mathbb{Z}$, we have $K = \mathbb{Q}(\sqrt{r}) = \mathbb{Q}(\sqrt{n})$.

- We now finish by proving that there exists a square-free $d \in \mathbb{Z}$ with $K = \mathbb{Q}(\sqrt{d})$. Fix $n \in \mathbb{Z}$ with $K = \mathbb{Q}(\sqrt{n})$. Notice that $n \neq 0$ because $K \neq \mathbb{Q}$. By the above lemma, we can write $n = m^2 d$ where $m \in \mathbb{N}^+$ and $d \in \mathbb{Z}$ is square-free. We then have $\sqrt{n} = m \cdot \sqrt{d}$ because $m > 0$. Thus, $\sqrt{n} \in \mathbb{Q}(\sqrt{d})$ and $\sqrt{d} = \frac{1}{m}\sqrt{n} \in \mathbb{Q}(\sqrt{n})$. It follows that $K = \mathbb{Q}(\sqrt{n}) = \mathbb{Q}(\sqrt{d})$.

This completes the proof. □

Theorem 5.1.5. *If $c, d \in \mathbb{Z}$ are both square-free and $\mathbb{Q}(\sqrt{c}) = \mathbb{Q}(\sqrt{d})$, then $c = d$.*

Proof. Suppose that $c, d \in \mathbb{Z}$ are both square-free and that $\mathbb{Q}(\sqrt{c}) = \mathbb{Q}(\sqrt{d})$. We then have $\sqrt{c} \in \mathbb{Q}(\sqrt{d})$. We know that $[\mathbb{Q}(\sqrt{d}) : \mathbb{Q}] = 2$, so

$$\mathbb{Q}(\sqrt{d}) = \{s + t\sqrt{d} : a, b \in \mathbb{Q}\}$$

Fix $s, t \in \mathbb{Q}$ with $\sqrt{c} = s + t\sqrt{d}$. Notice first that $t \neq 0$ because $\sqrt{c} \notin \mathbb{Q}$ as c is square-free (see the proof of the previous theorem).

Suppose that $s \neq 0$. Squaring both sides we see that

$$c = s^2 + 2st\sqrt{d} + dt^2$$

hence

$$\sqrt{d} = \frac{c - s^2 - dt^2}{2st} \in \mathbb{Q}$$

which is a contradiction.

Therefore, we must have $s = 0$ and $t \neq 0$. Let $t = \frac{a}{b}$ where $a, b \in \mathbb{Z}$ are relatively prime and $b > 0$. We have that $\sqrt{c} = t\sqrt{d}$. Squaring both sides gives $c = dt^2$ and multiplying both sides by b^2 we conclude that $b^2 c = a^2 d$. By the uniqueness part of Lemma 5.1.3 (applied to $-a$ and b if $a < 0$), we must have $c = d$. □

5.2 Integers in Quadratic Number Fields

Recall that given a number field K, we defined $\mathcal{O}_K$ to be the ring of algebraic integers in K. Now that we classified all quadratic number fields as $K = \mathbb{Q}(\sqrt{d})$ where $d \in \mathbb{Z}$ is square-free and $d \neq 1$, we set about the task of finding the ring of algebraic integers $\mathcal{O}_K$ in each of these number fields. As alluded to above, it is natural to believe that when $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ might be $\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{-d} : a, b \in \mathbb{Z}$. However, consider the case where $d = -3$. Letting $\zeta_3 = e^{2\pi i/3}$, we showed above that $\mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta_3)$. Now

$$\zeta_3 = -\frac{1}{2} + \frac{1}{2} \cdot \sqrt{-3}$$

and ζ_3 is a root of $x^3 - 1 = (x-1)(x^2 + x + 1)$. Thus, ζ is a root of $x^2 + x + 1 \in \mathbb{Z}[x]$, and hence ζ is an algebraic integer in $\mathbb{Q}(\sqrt{-3})$ that is not an element of $\mathbb{Z}[\sqrt{-3}]$. Before precisely determining $\mathcal{O}_K$ for the quadratic number fields K, we first prove a very useful lemma.

Lemma 5.2.1. *Let $d \in \mathbb{Z}$ be square-free. Suppose that $q \in \mathbb{Q}$ and $q^2 d \in \mathbb{Z}$. We then have that $q \in \mathbb{Z}$.*

Proof. Write $q = \frac{a}{b}$ where $a, b \in \mathbb{Z}$ and $b \neq 0$. Notice that

$$a^2 d = b^2 \cdot \left(\frac{a}{b}\right)^2 \cdot d = b^2 \cdot (q^2 d)$$

Since $q^2 d \in \mathbb{Z}$, we see that $b^2 \mid a^2 d$ in $\mathbb{Z}$. Let $p \in \mathbb{N}^+$ be an arbitrary prime. Since $b^2 \mid a^2 d$, we know that

$$ord_p(b^2) \leq ord_p(a^2 d)$$

and hence

$$2 \cdot ord_p(b) \leq 2 \cdot ord_p(a) + ord_p(d)$$

Now d is square-free, so we know that $ord_p(d) \leq 1$. Since $2 \cdot ord_p(b)$ and $2 \cdot ord_p(a)$ are both even and $ord_p(d) \in \{0, 1\}$, we can conclude that $2 \cdot ord_p(b) \leq 2 \cdot ord_p(a)$ and hence $ord_p(b) \leq ord_p(a)$. Since p was an arbitrary prime, it follows that $b \mid a$ in $\mathbb{Z}$. Therefore, $q = \frac{a}{b} \in \mathbb{Z}$. □

Theorem 5.2.2. *Suppose that $d \in \mathbb{Z}$ is square-free with $d \neq 1$.*

- *If $d \not\equiv 1 \pmod 4$, then the set of algebraic integers in the number field $\mathbb{Q}(\sqrt{d})$ is the set*

$$\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}$$

- *If $d \equiv 1 \pmod 4$, then the set of algebraic integers in the number field $\mathbb{Q}(\sqrt{d})$ is the set*

$$\begin{aligned} \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] &= \left\{a + b\left(\frac{1+\sqrt{d}}{2}\right) : a, b \in \mathbb{Z}\right\} \\ &= \{a + b\sqrt{d} : a, b \in \mathbb{Z}\} \cup \left\{\frac{a}{2} + \frac{b}{2}\sqrt{d} : a, b \in \mathbb{Z} \text{ are both odd}\right\} \end{aligned}$$

Proof. We first show that

$$\left\{a + b\left(\frac{1+\sqrt{d}}{2}\right) : a, b \in \mathbb{Z}\right\} = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\} \cup \left\{\frac{a}{2} + \frac{b}{2}\sqrt{d} : a, b \in \mathbb{Z} \text{ are both odd}\right\}$$

Let $a, b \in \mathbb{Z}$. We have

$$a + b\left(\frac{1+\sqrt{d}}{2}\right) = \left(a + \frac{b}{2}\right) + \frac{b}{2} \cdot \sqrt{d} = \left(\frac{2a+b}{2}\right) + \frac{b}{2} \cdot \sqrt{d}$$

If b is even, then $a + \frac{b}{2}$ and $\frac{b}{2}$ are integers, so this element is in the set on the right. If b is odd, then $2a + b$ is also odd, so again this element is in the set on the right. It follows that the set on the left is a subset of the one on the right. We now show the reverse containment. If $a, b \in \mathbb{Z}$, then

$$a + b\sqrt{d} = (a - b) + 2b \cdot \left(\frac{1 + \sqrt{d}}{2}\right)$$

Furthermore, if $a, b \in \mathbb{Z}$ are both odd, then $a - b$ is even and we have

$$\frac{a}{2} + \frac{b}{2}\sqrt{d} = \frac{a - b}{2} + b \cdot \left(\frac{1 + \sqrt{d}}{2}\right)$$

Therefore, the set on the right is a subset of the one on the left, and combining this with the above we conclude that the two sets are equal.

We now prove the result. Let $s, t \in \mathbb{Q}$ so $s + t\sqrt{d} \in \mathbb{Q}(\sqrt{d})$. Notice that

$$(s + t\sqrt{d})^2 = s^2 + 2st\sqrt{d} + dt^2 = (s^2 + dt^2) + (2st)\sqrt{d}$$

Therefore, we have

$$\begin{aligned}(s + t\sqrt{d})^2 - 2s \cdot (s + t\sqrt{d}) &= s^2 + 2st\sqrt{d} + dt^2 - 2s^2 - 2st\sqrt{d} \\ &= -s^2 + dt^2\end{aligned}$$

It follows that $s + t\sqrt{d}$ is root of the monic polynomial

$$x^2 + (-2s)x + (s^2 - dt^2)$$

Now if $s, t \in \mathbb{Z}$, then this polynomial is a monic polynomial with integer coefficients having $s + t\sqrt{d}$ as a root, so $s + t\sqrt{d}$ is an algebraic integer. In other words, every element in $\mathbb{Z}[\sqrt{d}]$ is an algebraic integer. Suppose that $d \equiv 1 \pmod 4$ and $s = \frac{a}{2}$ and $t = \frac{b}{2}$ where $a, b \in \mathbb{Z}$ are odd. Notice that $-2s = -a \in \mathbb{Z}$. We have

$$s^2 - dt^2 = \frac{a^2}{4} - d \cdot \frac{b^2}{4} = \frac{a^2 - db^2}{4}$$

Now since both a and b are odd, it follows that $a^2 \equiv 1 \pmod 4$ and $b^2 \equiv 1 \pmod 4$. Thus, $a^2 - db^2 \equiv (1 - 1 \cdot 1) \equiv 0 \pmod 4$, hence $4 \mid (a^2 - db^2)$. It follows that $s^2 - dt^2 \in \mathbb{Z}$ also, so $s + t\sqrt{d}$ is an algebraic integer. Therefore, if $d \equiv 1 \pmod 4$, then every element in $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ is an algebraic integer. We have shown that if $d \not\equiv 1 \pmod 4$, then every element of $\mathbb{Z}[\sqrt{d}]$ is an algebraic integer, and if $d \equiv 1 \pmod 4$, then every element of $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ is an algebraic integer.

We now show the converse. Suppose that $s, t \in \mathbb{Q}$ and $s + t\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ is an algebraic integer. If $t = 0$, then $s + t\sqrt{d} = s \in \mathbb{Q}$, so $s \in \mathbb{Z}$ (because $\mathbb{Z}$ is the set of algebraic integers in $\mathbb{Q}$) and hence $s + t\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ (and so $s + t\sqrt{d} \in \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ if $d \equiv 1 \pmod 4$). Suppose then that $t \neq 0$. We know that $s + t\sqrt{d}$ is a root of polynomial

$$x^2 + (-2s)x + (s^2 - dt^2)$$

and since this degree two polynomial has no rational roots (the roots are $s + t\sqrt{d}$ and $s - t\sqrt{d}$ and neither is rational because $t \neq 0$ and d is squarefree), it follows that it is irreducible and hence the minimal polynomial of $s + t\sqrt{d}$ over $\mathbb{Q}$. Since we are assuming that $s + t\sqrt{d}$ is an algebraic integer, we conclude from above that both $-2s \in \mathbb{Z}$ and $s^2 - dt^2 \in \mathbb{Z}$. We now have two cases.

- Suppose that $s \in \mathbb{Z}$. We then have that $s^2 \in \mathbb{Z}$, so since $s^2 - dt^2 \in \mathbb{Z}$, it follows that $dt^2 = s^2 - (s^2 - dt^2) \in \mathbb{Z}$. Since d is square-free, we may use the above lemma to conclude that $t \in \mathbb{Z}$. Therefore, $s + t\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$.
- Suppose that $s \notin \mathbb{Z}$. We have $-2s \in \mathbb{Z}$, so $2s \in \mathbb{Z}$, and hence we may write $s = \frac{a}{2}$ for some odd $a \in \mathbb{Z}$. Now
$$s^2 - td^2 = \frac{a^2}{4} - td^2 \in \mathbb{Z}$$
so $a^2 - 4td^2 \in \mathbb{Z}$, and hence $d(2t)^2 = 4dt^2 \in \mathbb{Z}$. Since d is square-free, the lemma implies that $2t \in \mathbb{Z}$. Fix $b \in \mathbb{Z}$ with $t = \frac{b}{2}$. We now have
$$s + t\sqrt{d} = \frac{a}{2} + \frac{b}{2}\sqrt{d}$$
with a odd. We know that
$$\frac{a^2 - b^2 d}{4} = s^2 - dt^2 \in \mathbb{Z}$$
and hence $4 \mid (a^2 - db^2)$. Now if b is even, then $2 \mid db^2$, so $2 \mid a^2$, a contradiction. Therefore, b is odd. The only thing left to do is to show that in this case we must have $d \equiv 1 \pmod 4$. Since a and b are both odd, we know that $a^2 \equiv 1 \equiv b^2 \pmod 4$. As noted above, we have $4 \mid (a^2 - b^2 d)$, so $a^2 \equiv b^2 d \pmod 4$. Therefore, $1 \equiv d \pmod 4$.

□

Notice in the special case where $d = -1$, we have $-1 \not\equiv 1 \pmod 4$, so
$$\mathcal{O}_{\mathbb{Q}(i)} = \{a + bi : a, b \in \mathbb{Z}\} = \mathbb{Z}[i]$$
as we probably expected.

Consider the case where $d = -3$. Recall that if we let
$$\zeta_3 = e^{2\pi i/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$$
then the minimal polynomial of ζ_3 over $\mathbb{Q}$ is $x^2 + x + 1$ and we have $\mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta_3)$. Consider $\zeta_6 = e^{2\pi i/6} = e^{\pi i/6}$. Now ζ_6 is a root of the polynomial $x^6 - 1$ and
$$x^6 - 1 = (x^3 - 1)(x^3 + 1) = (x - 1)(x + 1)(x^2 + x + 1)(x^2 - x + 1)$$
Thus, ζ_6 is a root of one of the four factors on the right, and a simple check shows that it is a root of $x^2 - x + 1$. This polynomial is irreducible over $\mathbb{Q}$ (because it has no rational roots), so it is the minimal polynomial of ζ_6 over $\mathbb{Q}$. Now
$$\zeta_6 = \cos(\pi/6) + i\sin(\pi/6) = \frac{1}{2} + \frac{\sqrt{3}}{2}i$$
It follows that $\zeta_6 = \zeta_3 + 1$, hence $\zeta_6 \in \mathbb{Q}(\zeta_3)$ and $\zeta_3 = \zeta_6 - 1 \in \mathbb{Q}(\zeta_6)$. We conclue that
$$\mathbb{Q}(\zeta_6) = \mathbb{Q}(\zeta_3) = \mathbb{Q}(\sqrt{-3})$$
Now $-3 \equiv 1 \pmod 4$, so we know that
$$\mathcal{O}_{\mathbb{Q}(\sqrt{-3})} = \mathbb{Z}\left[\frac{1 + \sqrt{-3}}{2}\right] = \mathbb{Z}[\zeta_6]$$
Since $\zeta_6 = \zeta_3 + 1$, we have $\zeta_6 \in \mathbb{Z}[\zeta_3]$ and $\zeta_3 \in \mathbb{Z}[\zeta_6]$, so we also have
$$\mathcal{O}_{\mathbb{Q}(\sqrt{-3})} = \mathbb{Z}[\zeta_3]$$
Thus, the ring of integers in the number field $\mathbb{Q}(\sqrt{-3})$ can be written in either of the following ways:
$$\mathcal{O}_{\mathbb{Q}(\sqrt{-3})} = \mathbb{Z}[\zeta_3] = \{a + b\zeta_3 : a, b \in \mathbb{Z}\} \qquad \mathcal{O}_{\mathbb{Q}(\sqrt{-3})} = \mathbb{Z}[\zeta_6] = \{a + b\zeta_6 : a, b \in \mathbb{Z}\}$$

5.3 Norms and Units

5.3.1 The Norm on a Quadratic Number Field

Definition 5.3.1. *Let $d \in \mathbb{Z}$ be square-free with $d \neq 1$. We define a function $N\colon \mathbb{Q}(\sqrt{d}) \to \mathbb{Q}$ by letting $N(s+t\sqrt{d}) = s^2 - dt^2$. The function N is called the* norm *on the field $\mathbb{Q}(\sqrt{d})$.*

Notice that if $d = -1$ and we are working in $\mathbb{Q}(\sqrt{-1}) = \mathbb{Q}(i)$, then

$$N(s+ti) = s^2 - (-1)t^2 = s^2 + t^2$$

is our old norm function.

Suppose that $d \in \mathbb{Z}\backslash\{1\}$ is square-free and $s, t \in \mathbb{Q}$. We showed in the above proof that $s+t\sqrt{d}$ is a root of the polynomial

$$x^2 + (-2s)x + (s^2 - dt^2)$$

As described in the above proof, this polynomial is irreducible in $\mathbb{Q}[x]$ when $t \neq 0$, and hence is the minimal polynomial of $s+t\sqrt{d}$ over $\mathbb{Q}$ when $t \neq 0$. Using the quadratic formula, the roots of $x^2 + (-2s)x + (s^2 - dt^2)$ are

$$\begin{aligned}\frac{-(-2s) \pm \sqrt{(-2s)^2 - 4(s^2 - dt^2)}}{2} &= \frac{2s \pm \sqrt{4dt^2}}{2} \\ &= \frac{2s \pm 2t\sqrt{d}}{2} \\ &= s \pm t\sqrt{d}\end{aligned}$$

With this in mind, we can interpret $N(s+t\sqrt{d})$ in a few ways when $t \neq 0$. One way is that $N(s+t\sqrt{d})$ is the constant term of the minimal polynomial of $s+t\sqrt{d}$ over $\mathbb{Q}$. Alternatively, $N(s+t\sqrt{d})$ is the product of the two roots of the minimal polynomial of $s+t\sqrt{d}$ over $\mathbb{Q}$. This is true because

$$N(s+t\sqrt{d}) = s^2 - dt^2 = (s+t\sqrt{d})(s-t\sqrt{d})$$

even in the case when $t = 0$. In the case where $d < 0$, notice that $N(s+t\sqrt{d}) = s^2 - dt^2$ is just the square of the distance between the complex point $s+t\sqrt{d} = s + (-t\sqrt{-d})i$ and the origin in the complex plane.

If you know some Galois theory, then $N(s+t\sqrt{d})$ is just the product of the two Galois conjugates of $s+t\sqrt{d}$ over $\mathbb{Q}$. If you don't know some Galois theory, then we elaborate on this idea now.

Definition 5.3.2. *Let $d \in \mathbb{Z}\backslash\{1\}$ be square-free. Define a function $\phi\colon \mathbb{Q}(\sqrt{d}) \to \mathbb{Q}(\sqrt{d})$ by letting*

$$\phi(s+t\sqrt{d}) = s - t\sqrt{d}$$

Given $\alpha \in \mathbb{Q}(\sqrt{d})$, we call $\phi(\alpha)$ the conjugate *of α, and denote it by $\overline{\alpha}$.*

Notice that if $d < 0$, then $\overline{\alpha}$ is indeed the normal complex conjugate of α. However, if $d > 0$, then $\mathbb{Q}(\sqrt{d}) \subseteq \mathbb{R}$ and hence $\overline{\alpha}$ is *not* the complex conjugate of α. For example, if $d = 2$ and we are working in $\mathbb{Q}(\sqrt{2})$, then $\overline{5+\sqrt{2}} = 5 - \sqrt{2}$ which are distinct positive real numbers.

Theorem 5.3.3. *Let $d \in \mathbb{Z}\backslash\{1\}$ be square-free and let $K = \mathbb{Q}(\sqrt{d})$. The conjugation map $\phi\colon K \to K$ defined above by $\phi(s+t\sqrt{d}) = s - t\sqrt{d}$ is a automorphism of the field K. Furthermore, ϕ maps $\mathcal{O}_K$ onto $\mathcal{O}_K$, so upon restriction to this subring we can view ϕ as automorphism of the ring of integers of K.*

Proof. Given $q, r, s, t \in \mathbb{Q}$, we have

$$\begin{aligned}\phi((q + r\sqrt{d}) + (s + t\sqrt{d})) &= \phi((q + s) + (r + t)\sqrt{d}) \\ &= (q + s) - (r + t)\sqrt{d} \\ &= (q - r\sqrt{d}) + (s - t\sqrt{d}) \\ &= \phi(q + r\sqrt{d}) + \phi(s + t\sqrt{d})\end{aligned}$$

and also

$$\begin{aligned}\phi((q + r\sqrt{d}) \cdot (s + t\sqrt{d})) &= \phi(qs + rs\sqrt{d} + qt\sqrt{d} + rtd) \\ &= \phi((qs + rtd) + (rs + qt)\sqrt{d}) \\ &= (qs + rtd) - (rs + qt)\sqrt{d} \\ &= qs - rs\sqrt{d} - qt\sqrt{d} + rtd \\ &= (q - r\sqrt{d}) \cdot (s - t\sqrt{d}) \\ &= \phi(q + r\sqrt{d}) \cdot \phi(s + t\sqrt{d})\end{aligned}$$

We also have $\phi(1) = \phi(1 + 0\sqrt{d}) = 1 - 0\sqrt{d} = 1$, so ϕ is an ring homomorphism. Now for any $q, r \in \mathbb{Q}$, we have

$$\begin{aligned}\phi(\phi(q + r\sqrt{d})) &= \phi(q - r\sqrt{d}) \\ &= \phi(q + (-r)\sqrt{d}) \\ &= q - (-r)\sqrt{d} \\ &= q + r\sqrt{d}\end{aligned}$$

Thus, $\phi^2 = id_K$, so ϕ is its own inverse and hence ϕ is a bijection. Therefore, ϕ is an automorphism of K.

We now prove that ϕ maps $\mathcal{O}_K$ *into* $\mathcal{O}_K$. One approach is simply to use our characterization of $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ to check it but we give another much more general argument. Suppose that $\alpha \in \mathcal{O}_K$. We then have that α is an algebraic integer, so we may fix a monic $h(x) \in \mathbb{Z}[x]$ with $h(\alpha) = 0$. Write

$$h(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

where each $a_i \in \mathbb{Z}$. Since $h(\alpha) = 0$, we have

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$$

Applying ϕ to both sides and using the fact that it preserves addition/multiplication, we conclude that

$$(\phi(\alpha))^n + a_{n-1}(\phi(\alpha))^{n-1} + \cdots + a_1\phi(\alpha) + a_0 = 0$$

so $\phi(\alpha)$ is also a root of $h(x)$. Since $h(x)$ is a monic polynomial in $\mathbb{Z}[x]$, we conclude that $\phi(\alpha)$ is also an algebraic integer. Since $\phi(\alpha) \in K$, we conclude that $\phi(\alpha) \in \mathcal{O}_K$. Therefore, ϕ maps $\mathcal{O}_K$ into $\mathcal{O}_K$. To finish the proof, we need to argue that every element of $\mathcal{O}_K$ is in the range of $\phi|_{\mathcal{O}_K}$. Let $\alpha \in \mathcal{O}_K$. We then have that $\phi(\alpha) \in \mathcal{O}_K$ by what we just proved and $\phi(\phi(\alpha)) = \alpha$. It follows that $\text{range}(\phi|_{\mathcal{O}_K}) = \mathcal{O}_K$. □

Using the above notation, notice that given $\alpha \in \mathbb{Q}(\sqrt{d})$, we have $N(\alpha) = \alpha\overline{\alpha}$. In particular, if $\alpha \neq 0$, then

$$\alpha^{-1} = \frac{\overline{\alpha}}{N(\alpha)}$$

Proposition 5.3.4. *Let $d \in \mathbb{Z}$ be squarefree and let $K = \mathbb{Q}(\sqrt{d})$. For the function $N(s+t\sqrt{d}) = s^2 - dt^2$ defined on $\mathbb{Q}(\sqrt{d})$, we have*

1. *If $d < 0$, then $N(\alpha) \geq 0$ for all $\alpha \in K$.*
2. *$N(\alpha) = 0$ if and only if $\alpha = 0$.*
3. *$N(q) = q^2$ for all $q \in \mathbb{Q}$.*
4. *$N(\alpha) \in \mathbb{Z}$ for all $\alpha \in \mathcal{O}_K$.*
5. *$N(\alpha\beta) = N(\alpha) \cdot N(\beta)$ for all $\alpha, \beta \in K$.*

Proof. Statements 1 and 3 are immediate from the definition. Statement 4 follows from the fact that $N(s+t\sqrt{d})$ is the constant term of the minimal polynomial of $s+t\sqrt{d}$ when $t \neq 0$, which is an integer because we are assuming $s + t\sqrt{d} \in \mathcal{O}_K$. For statement 2, we clearly have that $N(0) = 0$. Suppose that $N(\alpha) = 0$. Writing $\alpha = s + t\sqrt{d}$ where $s, t \in \mathbb{Q}$, we see that

$$(s+t\sqrt{d})(s-t\sqrt{d}) = N(\sqrt{d}) = 0$$

so either $s + t\sqrt{d} = 0$ or $s - t\sqrt{d} = 0$. Since $\{1, \sqrt{d}\}$ is a basis of $\mathbb{Q}(\sqrt{d})$ over $\mathbb{Q}$, it follows that $s = t = 0$, so $\alpha = 0$.

For statement 5, suppose that $\alpha, \beta \in \mathbb{Q}(\sqrt{d})$ and write $\alpha = q + r\sqrt{d}$ and $\beta = s + t\sqrt{d}$. We have

$$\begin{aligned} N(\alpha\beta) &= (\alpha\beta) \cdot \overline{(\alpha\beta)} \\ &= \alpha\beta\overline{\alpha}\,\overline{\beta} \\ &= \alpha\overline{\alpha}\beta\overline{\beta} \\ &= N(\alpha) \cdot N(\beta) \end{aligned}$$

□

Proposition 5.3.5. *Let $d \in \mathbb{Z}\backslash\{1\}$ be square-free, let $K = \mathbb{Q}(\sqrt{d})$, and let $\alpha \in \mathcal{O}_K$. We have that $\alpha \in U(\mathcal{O}_K)$ if and only if $N(\alpha) = \pm 1$.*

Proof. Suppose first that $\alpha \in U(\mathcal{O}_K)$. Fix $\beta \in \mathcal{O}_K$ with $\alpha\beta = 1$. We then have

$$N(\alpha)N(\beta) = N(\alpha\beta) = N(1) = 1^2 = 1$$

Since $N(\alpha), N(\beta) \in \mathbb{Z}$, it follows that $N(\alpha) \mid 1$ in $\mathbb{Z}$, so $N(\alpha) = \pm 1$.

Suppose conversely that $\alpha \in \mathcal{O}_K$ with $N(\alpha) = \pm 1$. If $N(\alpha) = 1$, then

$$1 = N(\alpha) = \alpha\overline{\alpha}$$

Since $\overline{\alpha} = \phi(\alpha) \in \mathcal{O}_K$, it follows that α has an inverse in $\mathcal{O}_K$, so is unit. Similarly, if $N(\alpha) = -1$, then

$$-1 = N(\alpha) = \alpha\overline{\alpha}$$

so $1 = \alpha\dot{(}-\overline{\alpha})$. Since $-\overline{\alpha} = -\phi(\alpha) \in \mathcal{O}_K$ (recall that $\mathcal{O}_K$ is a subring of K), it follows that $\alpha \in U(\mathcal{O}_K)$ □

Proposition 5.3.6. *Suppose that $d \in \mathbb{Z}$ is squarefree and that $d < 0$. Let $K = \mathbb{Q}(\sqrt{d})$.*

- *If $d = -1$, then $U(\mathcal{O}_K) = \{1, -1, i, -i\}$.*

- *If $d = -3$, then*

$$\begin{aligned} U(\mathcal{O}_K) &= \left\{1, -1, \frac{1}{2} + \frac{\sqrt{-3}}{2}, -\frac{1}{2} + \frac{\sqrt{-3}}{2}, -\frac{1}{2} - \frac{\sqrt{-3}}{2}, \frac{1}{2} - \frac{\sqrt{-3}}{2}\right\} \\ &= \{1, -1, \zeta_6, \zeta_6^2, \zeta_6^4, \zeta_6^5\} \\ &= \{1, -1, \zeta_6, \zeta_3, \zeta_3^{-1}, \zeta_6^{-1}\} \end{aligned}$$

- *If $d \notin \{-1, -3\}$, then $U(\mathcal{O}_K) = \{1, -1\}$.*

Proof. If $d = -1$, then $\mathcal{O}_K = \mathbb{Z}[i]$, and we've already seen that $U(\mathbb{Z}[i]) = \{1, -1, i, -i\}$. Suppose that $d \leq -2$. Let $m = -d$, so $m \in \mathbb{Z}$ and $m \geq 2$. Recall that $N(\alpha) \geq 0$ for all $\alpha \in \mathcal{O}_K$ because $d < 0$, and so $\alpha \in \mathcal{O}_K$ is a unit if and only if $N(\alpha) = 1$.

Suppose that $a, b \in \mathbb{Z}$. We have

$$N(a + b\sqrt{d}) = a^2 - db^2 = a^2 + mb^2$$

Thus, if $b \neq 0$, then $|b| \geq 1$, and hence $N(a + b\sqrt{d}) \geq mb^2 \geq m > 1$. Also, if $|a| \geq 2$, then $N(a + b\sqrt{d}) \geq a^2 \geq 4 > 1$. It follows that $N(a + b\sqrt{d}) = 1$ if and only if either $(a, b) = (1, 0)$ or $(a, b) = (-1, 0)$, so the only units of this form are 1 and -1.

Suppose now that $d \equiv 1 \pmod 4$. Let $a, b \in \mathbb{Z}$ with both a and b odd so that $\frac{a}{2} + \frac{b}{2}\sqrt{d} \in \mathcal{O}_K$. We have

$$N\left(\frac{a}{2} + \frac{b}{2}\sqrt{d}\right) = \left(\frac{a}{2}\right)^2 - d \cdot \left(\frac{b}{2}\right)^2 = \frac{a^2 - db^2}{4} = \frac{a^2 + mb^2}{4}$$

We therefore have that $N(\frac{a}{2} + \frac{b}{2}\sqrt{d}) = 1$ if and only if $a^2 + mb^2 = 4$. If $d \neq -3$, then $d \leq -7$, so $m \geq 7$ and hence when $b \neq 0$ we have $a^2 + mb^2 \geq m > 4$. Thus, if $d \neq -3$, then we are looking for solutions to $a^2 = 4$ in the odd integers, which do not exist. Suppose then that $d = -3$ so $m = 3$. We are now looking for solutions to $a^2 + 3b^2 = 4$ where $a, b \in \mathbb{Z}$ are odd. If $b = 0$, then $a^2 = 4$ which as above has no solutions in odd integers. If $|b| \geq 2$, then $a^2 + 3b^2 \geq 12 > 4$. If $b = \pm 1$, then we are looking for solutions to $a^2 + 3 = 4$ where $a \in \mathbb{Z}$ are odd, which clearly has solutions $a = \pm 1$. This gives the above units. □

5.3.2 Units in Real Quadratic Number Fields and Pell's Equation

When we examine $d > 0$, the situation is much more interesting. Consider the case where $d = 2$. We have $K = \mathbb{Q}(\sqrt{2})$ and $\mathcal{O}_K = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$. We know that

$$U(\mathcal{O}_K) = \{\alpha \in \mathcal{O}_K : N(\alpha) = \pm 1\}$$

and that

$$N(a + b\sqrt{2}) = a^2 - 2b^2$$

Thus, given $a, b \in \mathbb{Z}$, we have $a + b\sqrt{2} \in U(\mathcal{O}_K)$ if and only if $a^2 - 2b^2 = \pm 1$. One such example is $1 + \sqrt{2}$. Notice that

$$N(1 + \sqrt{2}) = 1^2 - 2 \cdot 1^2 = -1$$

so $1 + \sqrt{2} \in U(\mathcal{O}_K)$. We have

$$(1 + \sqrt{2})^{-1} = \frac{1 - \sqrt{2}}{-1} = -1 + \sqrt{2}$$

Now we know that $U(\mathcal{O}_K)$ is a multiplication subgroup of $\mathbb{O}_K$, so the product of two units is a unit. Thus, we can obtain more elements of $U(\mathcal{O}_K)$ by taking powers of $1 + \sqrt{2}$. Now since $1 + \sqrt{2} \in \mathbb{R}$ with $1 + \sqrt{2} > 1$,

these powers will increase in size in $\mathbb{Q}$, and hence will never repeat. Thus, $U(\mathcal{O}_K)$ will be infinite. For example, another element of $U(\mathcal{O}_K)$ is

$$(1+\sqrt{2})^2 = 1 + 2\sqrt{2} + 2 = 3 + 2\sqrt{2}$$

We have

$$N(3+2\sqrt{2}) = N((1+\sqrt{2})^2) = N(1+\sqrt{2})^2 = (-1)^2 = 1$$

and

$$(3+2\sqrt{2})^{-1} = 3 - \sqrt{2}$$

Multiplying by $1+\sqrt{2}$ again we obtain

$$(3+2\sqrt{2})(1+\sqrt{2}) = 3 + 3\sqrt{2} + 2\sqrt{2} + 4 = 7 + 5\sqrt{2}$$

which is another unit with norm -1. If we keep taking powers we obtain units whose norm alternates between 1 and -1. We have the following list:

$$\begin{aligned}
(1+\sqrt{2})^1 &= 1 + \sqrt{2} \\
(1+\sqrt{2})^2 &= 3 + 2\sqrt{2} \\
(1+\sqrt{2})^3 &= 7 + 5\sqrt{2} \\
(1+\sqrt{2})^4 &= 17 + 12\sqrt{2} \\
(1+\sqrt{2})^5 &= 41 + 29\sqrt{2} \\
(1+\sqrt{2})^6 &= 99 + 70\sqrt{2}
\end{aligned}$$

One can obtain a recurrence to help calculate these. If $n \in \mathbb{N}^+$ is such that

$$(1+\sqrt{2})^n = a + b\sqrt{2}$$

Then

$$\begin{aligned}
(1+\sqrt{2})^{n+1} &= (1+\sqrt{2})^n \cdot (1+\sqrt{2}) \\
&= (a+b\sqrt{2})(1+\sqrt{2}) \\
&= (a+2b) + (a+b)\sqrt{2}
\end{aligned}$$

Notice that if you take one of the above units $a + b\sqrt{2}$, then $\frac{a}{b}$ is a pretty good approximation to $\sqrt{2}$. We have $\sqrt{2} = 1.414213\ldots$ while

$$\begin{aligned}
\frac{1}{1} &= 1.000000\ldots \\
\frac{3}{2} &= 1.500000\ldots \\
\frac{7}{5} &= 1.400000\ldots \\
\frac{17}{12} &= 1.416666\ldots \\
\frac{41}{29} &= 1.413793\ldots \\
\frac{99}{70} &= 1.414285\ldots
\end{aligned}$$

The study of units in $\mathcal{O}_K$ where $K = \mathbb{Q}(\sqrt{d})$ and $d \in \mathbb{Z}$ is square-free and at least 2 is closely related to another problem in the history of number theory.

Definition 5.3.7. *Let $n \in \mathbb{N}^+$. Pell's Equation (relative to n) is the equation $x^2 - ny^2 = 1$, where we look for solutions $x, y \in \mathbb{Z}$.*

Given $n \in \mathbb{N}^+$, two solutions to the Pell Equation are trivial: namely $(1, 0)$ and $(-1, 0)$. Notice that if (x, y) is a solution to the Pell Equation, then so are $(\pm x, \pm y)$, so we may focus attention on $x, y \in \mathbb{N}$. Furthermore, there are clearly no such solutions with $x = 0$, and the trivial solutions are the only ones with $y = 0$. Thus, we may assume that $x, y \in \mathbb{N}^+$ when examining nontrivial solutions.

We restrict to the case where $n \geq 1$ because if $n \in \mathbb{Z}$ with $n \leq n$, then the trivial solutions are clearly the only solutions. Suppose that n is a perfect square, say $n = m^2$. In this case, Pell's Equation can be easily solved by factoring. We have

$$x^2 - ny^2 = x^2 - m^2y^2 = (x - my)(x + my)$$

Thus, if (x, y) is a solution to Pell's Equation, then

$$(x - my)(x + my) = 1$$

Since $x - my, x + my \in \mathbb{Z}$, it follows that $x - my = \pm 1$ and $x + my = \pm 1$. In the case where $x - my = 1$ and $x + my = 1$, we see that $2x = 2$, so $x = 1$ and hence $y = 0$. In the case where $x - my = -1$ and $x + my = -1$, we see that $2x = -2$, so $x = -1$ and $y = 0$. Therefore, if n is a perfect square, then the trivial solutions are the only solutions.

We will focus attention on the case of Pell's Equation $x^2 - dy^2 = 1$ where d is square-free and at least 2. Given a general $n \in \mathbb{N}^+$ that is not a perfect square, we can write $n = dm^2$ where $d \geq 2$ is square-free and $m \in \mathbb{N}^+$. The equation

$$x^2 - ny^2 = 1$$

is then the same thing as the equation

$$x^2 - d(my)^2 = 1$$

Thus, if we understand solutions to $x^2 - dy^2 = 1$, we could in theory translate those solutions where $m \mid y$ into solutions of the equation with n.

Suppose then that $d \in \mathbb{Z}$ is square-free with $d \geq 2$. Let $K = \mathbb{Q}(\sqrt{d})$. Notice that (x, y) is a solution to Pell's Equation, then $x^2 - dy^2 = 1$, so

$$N(x + y\sqrt{d}) = x^2 - dy^2 = 1$$

and hence $x + y\sqrt{d} \in U(\mathcal{O}_K)$. Thus, solutions to Pell's Equation give units in $\mathcal{O}_K$. The converse is not quite true for two reasons. First, an $\alpha \in U(\mathcal{O}_K)$ could be such that $N(\alpha) = -1$ rather than $N(\alpha) = 1$. However, notice that in this case we have $\alpha^2 \in U(\mathcal{O}_K)$ and $N(\alpha^2) = 1$, so we can build a unit in $\mathcal{O}_K$ from such an element. The other possibility is that we could have $d \equiv 1 \pmod 4$ and our unit could be of the form $\frac{a}{2} + \frac{b}{2}\sqrt{d}$ where $a, b \in \mathbb{Z}$ are odd, which does not give rise to a solution to Pell's Equation.

Regardless, if we can show that Pell's Equation always has a nontrivial solution for $d \geq 2$ square-free, then we will have shown that there is always a nontrivial unit in $\mathcal{O}_K$. We now go about proving this important theorem.

Theorem 5.3.8. *Suppose that $d \in \mathbb{Z}$ is square-free and $d \geq 2$. There exist $x, y \in \mathbb{Z}$ with $x \neq \pm 1$ such that $x^2 - dy^2 = 1$. In other words, there exists a nontrivial solution to Pell's Equation.*

Before jumping into the proof of the theorem, we develop some more refined intuition about what a nontrivial solution provides in terms of rational approximations to $\sqrt{d}$. Suppose then that $(x, y) \in \mathbb{N}^+$ is a nontrivial solution to the Pell Equation. Notice that

$$(x - y\sqrt{d})(x + y\sqrt{d}) = x^2 - dy^2 = 1$$

and hence

$$x - y\sqrt{d} = \frac{1}{x + y\sqrt{d}}$$

It follows that

$$\frac{x}{y} - \sqrt{d} = \frac{1}{y \cdot (x + y\sqrt{d})}$$

Since $x + y\sqrt{d} > y$ (as $x > 0$), it follows that

$$0 < \frac{x}{y} - \sqrt{d} < \frac{1}{y^2}$$

so in particular

$$\left|\frac{x}{y} - \sqrt{d}\right| < \frac{1}{y^2}$$

Intuitively, this is saying that $\frac{x}{y}$ is a good approximation to the irrational number $\sqrt{d}$ because its distance from $\sqrt{d}$ is $\frac{1}{y^2}$, which is much less than $\frac{1}{y}$ (Notice that it is simple to argue that given $y \in \mathbb{N}^+$, there exists $x \in \mathbb{N}^+$ with $|\frac{x}{y} - \sqrt{d}| < \frac{1}{y}$). As we will see, solutions to the Pell's Equation and "good" rational approximations to $\sqrt{d}$ go hand-in-hand. In fact, we will prove the above theorem by proving a sequence of lemmas which slowly reverse the above implications. In particular, our first goal is to show that $\sqrt{d}$ has infinitely many "good" rational approximations.

Lemma 5.3.9. *Let $\alpha \in \mathbb{R}$ with $\alpha > 0$ and let $M \in \mathbb{N}^+$. There exists $x, y \in \mathbb{N}$ such that $0 < y \leq M$ and $|x - y\alpha| < \frac{1}{M}$.*

Proof. Divide the interval $[0, 1)$ into M subintervals

$$[0, \tfrac{1}{M}) \quad [\tfrac{1}{M}, \tfrac{2}{M}) \quad [\tfrac{2}{M}, \tfrac{3}{M}) \quad \cdots \quad [\tfrac{M-1}{M}, 1)$$

Consider the following $M + 1$ many elements of $[0, 1)$:

$$0\alpha - \lfloor 0\alpha \rfloor \quad 1\alpha - \lfloor 1\alpha \rfloor \quad 2\alpha - \lfloor 2\alpha \rfloor \quad \cdots \quad M\alpha - \lfloor M\alpha \rfloor$$

By the Pigeonhole Principle, two of these numbers lie in the same interval. Thus, there exists k, ℓ with $0 \leq k < \ell \leq M$ such that

$$|(k\alpha - \lfloor k\alpha \rfloor) - (\ell\alpha - \lfloor \ell\alpha \rfloor)| < \frac{1}{M}$$

We then have

$$|(\lfloor \ell\alpha \rfloor - \lfloor k\alpha \rfloor) - (\ell - k)\alpha| < \frac{1}{M}$$

Thus, we may let

$$x = \lfloor \ell\alpha \rfloor - \lfloor k\alpha \rfloor$$

and

$$y = \ell - k$$

Notice that $x \geq 0$ because $\ell\alpha > k\alpha$ (as $k < \ell$ and $\alpha > 0$) and also $0 < y \leq M$ because $1 \leq k < \ell \leq M + 1$. □

Lemma 5.3.10. *Let $\alpha \in \mathbb{R}$ be irrational with $\alpha > 0$. There exist infinitely many pairs $(x, y) \in \mathbb{N}^2$ with $y \neq 0$ such that*

$$\left|\frac{x}{y} - \alpha\right| < \frac{1}{y^2}$$

Proof. Notice that at least one such pair exists, namely $(\lfloor \alpha \rfloor, 1)$. Suppose that there exist n such pairs

$$(x_1, y_1) \quad (x_2, y_2) \quad (x_3, y_3) \quad \cdots \quad (x_n, y_n)$$

We show how to find another such pair. Let

$$\delta = \min\{|x_i - \alpha y_i| : 1 \leq i \leq n\}$$

and notice that $\delta > 0$ because α is irrational. Let $M \in \mathbb{N}^+$ be chosen such that $M > \frac{1}{\delta}$. By the previous lemma, there exist $x, y \in \mathbb{N}$ with $0 < y \leq M$ and $|x - y\alpha| < \frac{1}{M}$. Notice that $\frac{1}{M} < \delta$, so (x, y) is distinct from each of the pairs (x_i, y_i). Now $0 < y \leq M$, so

$$\left| \frac{x}{y} - \alpha \right| < \frac{1}{yM} \leq \frac{1}{y^2}$$

□

With the previous lemma in hand, we are now ready to prove that the Pell Equation has a nontrivial solution whenever $d \in \mathbb{N}^+$ is squarefree.

Proof of Theorem 5.3.8. Since $d \in \mathbb{Z}$ is square-free with $d \geq 2$, we know that $\sqrt{d}$ is irrational. For any $(x, y) \in \mathbb{N}^2$ with $y \neq 0$ and

$$\left| \frac{x}{y} - \alpha \right| < \frac{1}{y^2}$$

we can view $x + y\sqrt{d} \in \mathbb{Z}[\sqrt{d}] \subseteq \mathcal{O}_K$, where we have

$$\begin{aligned}
|N(x + y\sqrt{d})| &= |x^2 - dy^2| \\
&= |x - y\sqrt{d}| \cdot |x + y\sqrt{d}| \\
&= y^2 \cdot \left| \frac{x}{y} - \sqrt{d} \right| \cdot \left| \frac{x}{y} + \sqrt{d} \right| \\
&< y^2 \cdot \frac{1}{y^2} \cdot \left| \frac{x}{y} + \sqrt{d} \right| \\
&= \left| \frac{x}{y} + \sqrt{d} \right| \\
&= \left| \frac{x}{y} - \sqrt{d} + 2\sqrt{d} \right| \\
&\leq \left| \frac{x}{y} - \sqrt{d} \right| + 2\sqrt{d} \\
&< \frac{1}{y^2} + 2\sqrt{d} \\
&\leq 2\sqrt{d} + 1
\end{aligned}$$

Our goal now is to find a nontrivial element of $\mathbb{Z}[\sqrt{d}]$ of norm 1 (it's possible that $\mathbb{Z}[\sqrt{d}] \subsetneq \mathcal{O}_K$, but we just work in this smaller subring in that case). Using the previous lemma combined with the above calculations, it follows that there are infinitely many elements $\alpha \in \mathbb{Z}[\sqrt{d}]$ with

$$-(1 + 2\sqrt{d}) < N(\alpha) < 1 + 2\sqrt{d}$$

In particular, there exists a $k \in \mathbb{Z}$ such that there are infinitely many $\alpha \in \mathbb{Z}[\sqrt{d}]$ with $N(\alpha) = k$. Notice that $k \neq 0$ because we know that $N(\alpha) = 0$ implies $\alpha = 0$. Looking at the coefficients $a, b \in \mathbb{Z}$ of these

$\alpha = a + b\sqrt{d}$, we see that modulo $|k|$ there are at most $|k|$ possible values of each, and hence there are at most k^2 possible pairs modulo $|k|$, Thus, since there are infinitely many $\alpha \in \mathbb{Z}[\sqrt{d}]$ with $N(\alpha) = k$, we may fix $\alpha, \beta \in \mathbb{Z}[\sqrt{d}]$ such that

- $N(\alpha) = k = N(\beta)$
- $k \mid (\alpha - \beta)$ in $\mathbb{Z}[\sqrt{d}]$ (i.e. the corresponding coefficients of α and β are pairwise congruent modulo k).
- $\alpha \neq \pm\beta$.

Now $\beta \in \mathbb{Z}[\sqrt{d}] \subseteq \mathbb{Q}(\sqrt{d})$ and $\beta \neq 0$, so $\beta^{-1} \in \mathbb{Q}(\sqrt{d})$. Notice that since

$$N(\beta)N(\beta^{-1}) = N(\beta\beta^{-1}) = N(1) = 1$$

it follows that

$$N(\beta^{-1}) = \frac{1}{N(\beta)} = \frac{1}{k}$$

and thus

$$N(\alpha\beta^{-1}) = N(\alpha)N(\beta^{-1}) = k \cdot \frac{1}{k} = 1$$

We have now found a nontrivial element $\alpha\beta^{-1} \in \mathbb{Q}(\sqrt{d})$ of norm 1 (notice that it does not equal ± 1 because $\alpha \neq \pm\beta$), and our goal now is to show that our element is in $\mathbb{Z}[\sqrt{d}]$. We have

$$\alpha\beta^{-1} = \alpha \cdot \frac{\overline{\beta}}{N(\beta)} = \frac{\alpha\overline{\beta}}{k}$$

To finish, we need only show that $k \mid \alpha\overline{\beta}$ in $\mathbb{Z}[\sqrt{d}]$. Notice that $\beta\overline{\beta} = N(\beta) = k$ so

$$\begin{aligned} \alpha\overline{\beta} &= \alpha\overline{\beta} - k + k \\ &= \alpha\overline{\beta} - \beta\overline{\beta} + k \\ &= (\alpha - \beta)\overline{\beta} - k \end{aligned}$$

Since we chose α and β so that $k \mid (\alpha - \beta)$, it follows that $k \mid \alpha\overline{\beta}$ in $\mathbb{Z}[\sqrt{d}]$. Therefore, $\alpha\beta^{-1} \in \mathbb{Z}[\sqrt{d}]$ is a nontrivial element of norm 1, so gives a nontrivial solution to Pell's Equation. □

Lemma 5.3.11. *Let $d \in \mathbb{Z}$ be square-free with $d \geq 2$, and let $\mu \in \mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ be a unit with $\mu \neq \pm 1$. The elements $\pm\mu$ and $\pm\overline{\mu}$ are four distinct units, and exactly one of them lies in each of the intervals $(-\infty, -1)$, $(-1, 0)$, $(0, 1)$ and $(1, \infty)$.*

Proof. We have

$$\mu^{-1} = \frac{\overline{\mu}}{N(\mu)} = \frac{\overline{\mu}}{\pm 1} = \pm\overline{\mu}$$

Since $N(1) = 1 = N(-1)$ and N is multiplicative, it follows that each of the four elements $\pm\mu$ and $\pm\overline{\mu}$ are units, and they are equal to the four units $\pm\mu$ and $\pm\mu^{-1}$. Notice by assumption that none of them are ± 1. If $\mu > 1$, then $0 < \mu^{-1} < 1$, so $-1 < -\mu^{-1} < 0$ and $-\mu < -1$. The other cases are similar. □

Lemma 5.3.12. *Let $d \in \mathbb{Z}$ be square-free with $d \geq 2$, and let $\mu \in \mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ be a unit with $\mu \neq \pm 1$.*

- *If $\mu = a + b\sqrt{d}$ where $a, b \in \mathbb{Z}$, then $\mu > 1$ if and only if both $a > 0$ and $b > 0$.*
- *If $d \equiv 1 \pmod 4$ and $\mu = \frac{a}{2} + \frac{b}{2}\sqrt{d}$ where $a, b \in \mathbb{Z}$ are both odd, then $\mu > 1$ if and only if both $a > 0$ and $b > 0$.*

Proof. Suppose first that $\mu = a + b\sqrt{d}$ where $a, b \in \mathbb{Z}$. If both $a, b > 0$, then clearly $a + b\sqrt{d} \geq 1 + \sqrt{d} > 1$. Suppose conversely that $a + b\sqrt{d} > 1$. By the previous lemma, the four elements

$$a + b\sqrt{d}, a + (-b)\sqrt{d}, (-a) + b\sqrt{d}, (-a) + (-b)\sqrt{d}$$

are all units and exactly one of them is greater than 1. Now the choice with both coefficients positive is clearly greater than the other three, so that must be the one which is greater than 1. Since we are assuming that $a + b\sqrt{d} > 1$, it follows that both $a > 0$ and $b > 0$.

Suppose now that $d \equiv 1 \pmod 4$ and $\mu = \frac{a}{2} + \frac{b}{2}\sqrt{d}$ where $a, b \in \mathbb{Z}$ are both odd. We then have $d \geq 5$. If both $a, b > 0$, then $a + b\sqrt{d} \geq \frac{1}{2} + \frac{1}{2}\sqrt{5} > 1$. The converse is the same argument as in the first part. □

Corollary 5.3.13. *Let $d \in \mathbb{Z}$ be square-free with $d \geq 2$, and let $M \in \mathbb{R}$ with $M > 1$. There are only finitely many units in $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ lying in the interval $(1, M]$.*

Proof. Notice that there are only finitely many positive integers with $a + b\sqrt{d} \leq 2M$ and use the previous Lemma. □

Corollary 5.3.14. *Let $d \in \mathbb{Z}$ be square-free with $d \geq 2$. There exists $\mu_0 \in U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})})$ with $\mu_0 > 1$ such that $\mu_0 \leq \mu$ whenever $\mu \in U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})})$ satisfies $\mu > 1$.*

Proof. We know that there exists a nontrivial solution (x, y) to Pell's Equation, and we may assume that $x, y > 0$. We then have that $x + y\sqrt{d} > 1$ is a nontrivial unit. By the previous Corollary, there are only finitely many units in the interval $(1, x + y\sqrt{d}]$, and since there is at least one we may choose a smallest such unit μ_0. □

Definition 5.3.15. *Let $d \in \mathbb{Z}$ be square-free with $d \geq 2$. The unique unit $\mu_0 \in U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})})$ satisfying the above corollary is called the* fundamental unit *of $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$.*

Theorem 5.3.16. *Let $d \in \mathbb{Z}$ be square-free with $d \geq 2$. Let μ_0 be the fundamental unit of $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$. We then have*

$$U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})}) = \{\pm\mu_0^n : n \in \mathbb{Z}\}$$

In particular, $U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})}) \cong \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$ where the term on the right is though of as an additive abelian group.

Proof. Suppose first that $\nu \in U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})})$ is an arbitrary unit with $\nu > 1$. Let $n \in \mathbb{N}^+$ be least such that $\nu \leq \mu_0^n$ (notice that such an n exists because $\mu_0, \nu > 1$). We then have

$$\mu_0^{n-1} < \nu \leq \mu_0^n$$

Multiplying through by $\mu_0^{-(n-1)} > 0$, we conclude that

$$1 < \nu\mu_0^{-(n-1)} \leq \mu_0$$

Since $\nu\mu^{-(n-1)}$ is a unit, we must have $\nu\mu_0^{-(n-1)} = \mu_0$ by choice of μ_0. It follows that $\nu = \mu_0^n$. Thus we have shown that every unit greater than 1 is a positive power of μ_0.

If ν is a unit and $0 < \nu < 1$, then ν^{-1} is a unit with $\nu^{-1} > 1$, so by what we just showed it follows that $\nu^{-1} = \mu^n$ for some $n > 0$. We then have $\nu = \mu^{-n}$. Now if $\nu < 0$, notice that $-\nu$ is a unit with $-\nu > 0$, so we finish by applying what we just showed. Thus

$$U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})}) = \{\pm\mu_0^n : n \in \mathbb{Z}\}$$

For the final claim, define a function $\psi\colon \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z}) \to U(\mathcal{O}_{\mathbb{Q}(\sqrt{d})})$ by letting

$$\psi(n, \overline{k}) = (-1)^k\mu_0^n$$

Notice that ψ is well-defined because $(-1)^k = 1$ when k is even and $(-1)^k = -1$ when k is odd. It is straightforward to check that ψ is a homomorphism of abelian groups (where the left-hand side is viewed additively and the right-hand side is viewed multiplicatively). Since $\mu_0^n > 1$ for all $n \in \mathbb{N}^+$, we have $\ker(\psi) = (0, \overline{0})$, so ψ is injective. We know that ψ is surjective from above, so ψ is an isomorphism. □

5.4 Factorizations

Let $K = \mathbb{Q}(\sqrt{-5})$. Notice that $-5 \not\equiv 1 \pmod 4$, so

$$\mathcal{O}_K = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$$

Working in $\mathcal{O}_K$, we have

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

so it looks like we have two distinct factorizations of 6. Let's check that these really are two different factorizations into irreducibles. Now $N(2) = 4$, so if $\alpha, \beta \in \mathcal{O}_K$ are such that $2 = \alpha\beta$, then

$$4 = N(2) = N(\alpha\beta) = N(\alpha)N(\beta)$$

Since $d < 0$, all norms are nonnegative, so $N(\alpha), N(\beta) \in \{1, 2, 4\}$. Notice that there are no elements of norm 2 because $a^2 + 5b^2 = 2$ has no integer solutions. It follows that either $N(\alpha) = 1$ or $N(\beta) = 1$, so one of α or β is a unit. Therefore, 2 irreducible in $\mathcal{O}_K$. Since $a^2 + 5b^2 = 3$ also has no integer solutions, the same argument shows that 3 is irreducible in $\mathcal{O}_K$ as well. Now notice that

$$N(1 + \sqrt{-5}) = 6 \qquad N(1 - \sqrt{-5}) = 6$$

As above, suppose that $1 + \sqrt{-5} = \alpha\beta$. We then have

$$6 = N(1 + \sqrt{-5}) = N(\alpha\beta) = N(\alpha)N(\beta)$$

Since we just saw that there are no elements of norm 2 or 3, either $N(\alpha) = 1$ or $N(\beta) = 1$. Thus, either α is a unit or β is a unit, and hence $1 + \sqrt{-5}$ is irreducible in $\mathcal{O}_K$. The argument that $1 - \sqrt{-5}$ is irreducible in $\mathcal{O}_K$ is identical. Finally, note that the only units in $\mathcal{O}_K$ are ± 1, so these really are two distinct factorizations.

To get more out of this example, we have that 2 is irreducible in $\mathcal{O}_K$ and that

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

hence $2 \mid (1 + \sqrt{-5})(1 - \sqrt{-5})$ in $\mathcal{O}_K$. Now if $2 \mid (a + b\sqrt{-5})$, say $2(c + d\sqrt{-5}) = a + b\sqrt{-5}$, then $2c = a$ and $2d = b$, hence we would have both $2 \mid a$ and $2 \mid b$ in $\mathbb{Z}$. It follows that $2 \nmid (1 + \sqrt{-5})$ and $2 \nmid (1 - \sqrt{-5})$ in $\mathcal{O}_K$, so 2 is not prime in $\mathcal{O}_K$. Thus, in the ring $\mathcal{O}_K$, the notions of irreducible and prime are distinct.

Before we jump into general theory about what can be salvaged in these rings, we first try to quarantine off the rings which already behave well by classifying a few of these rings which do happen to be Euclidean domains.

Lemma 5.4.1. *Let $d \in \mathbb{N}^+$ be squarefree and let $K = \mathbb{Q}(\sqrt{d})$. Suppose that for all $\alpha, \beta \in \mathcal{O}_K$ with $\beta \neq 0$, there exists $\gamma \in \mathcal{O}_K$ such that*

$$\left| N\left(\frac{\alpha}{\beta} - \gamma\right) \right| < 1$$

We then have that $|N|$ is a Euclidean function on $\mathcal{O}_K$.

Proof. Let $\alpha, \beta \in \mathcal{O}_K$ with $\beta \neq 0$. By assumption, we may fix $\gamma \in \mathcal{O}_K$ with

$$\left| N\left(\frac{\alpha}{\beta} - \gamma\right)\right| < 1$$

Let $\rho = \alpha - \beta\gamma \in \mathcal{O}_K$. Notice that $\alpha = \beta\gamma + \rho$ and

$$\begin{aligned} |N(\rho)| &= |N(\alpha - \beta\gamma)| \\ &= \left| N\left(\beta \cdot \left(\frac{\alpha}{\beta} - \gamma\right)\right)\right| \\ &= \left| N(\beta) \cdot N\left(\frac{\alpha}{\beta} - \gamma\right)\right| \\ &= |N(\beta)| \cdot \left| N\left(\frac{\alpha}{\beta} - \gamma\right)\right| \\ &< |N(\beta)| \end{aligned}$$

Thus, $|N|$ is a Euclidean function on $\mathcal{O}_K$. □

Lemma 5.4.2. *If $a, b \in \mathbb{R}$ with $a, b \geq 0$, then $|a - b| \leq \max\{a, b\}$.*

Proof. If $0 \leq b \leq a$, then $0 \leq a - b \leq a$, hence

$$|a - b| = a - b \leq a = \max\{a, b\}$$

On the other hand,if $0 \leq a \leq b$, then $-b \leq a - b \leq 0$, hence

$$|a - b| = -(a - b) = b - a \leq b = \max\{a, b\}$$

□

Theorem 5.4.3. *Let $d \in \{-2, -1, 2, 3\}$ and let $K = \mathbb{Q}(\sqrt{d})$. The function $|N|$ is a Euclidean function on $\mathcal{O}_K$.*

Proof. Suppose that $\alpha, \beta \in \mathcal{O}_K$ with $\beta \neq 0$. We have $\frac{\alpha}{\beta} \in K$, so we may write $\frac{\alpha}{\beta} = s + t\sqrt{d}$ for some $s, t \in \mathbb{Q}$. Fix integers $m, n \in \mathbb{Z}$ closest to $s, t \in \mathbb{Q}$ respectively, i.e. fix $m, n \in \mathbb{Z}$ so that $|s - m| \leq \frac{1}{2}$ and $|t - n| \leq \frac{1}{2}$. Let $\gamma = m + n\sqrt{d} \in \mathcal{O}_K$. We then have

$$\begin{aligned} \left| N\left(\frac{\alpha}{\beta} - \gamma\right)\right| &= |N((s + t\sqrt{d}) - (m + n\sqrt{d}))| \\ &= |N((s - m) + (t - n)\sqrt{d})| \\ &= |(s - m)^2 - d(t - n)^2| \end{aligned}$$

Now if $d \in \{-2, -1\}$, then

$$\begin{aligned} \left| N\left(\frac{\alpha}{\beta} - \gamma\right)\right| &= |(s - m)^2 - d(t - n)^2| \\ &\leq (s - m)^2 + |d| \cdot (t - n)^2 \\ &\leq \frac{1}{4} + \frac{|d|}{4} \\ &< 1 \end{aligned}$$

If $d \in \{2, 3\}$, then

$$\begin{aligned}\left|N\left(\frac{\alpha}{\beta} - \gamma\right)\right| &= |(s-m)^2 - d(t-n)^2| \\ &\leq \max\{(s-m)^2, d(t-n)^2\} \qquad \text{(by the previous lemma)} \\ &\leq \max\left\{\frac{1}{4}, \frac{d}{4}\right\} \\ &< 1\end{aligned}$$

Therefore, $|N|$ is a Euclidean function by the above lemma. □

Theorem 5.4.4. *Let $d \in \{-11, -7, -3, 5, 13\}$ and let $K = \mathbb{Q}(\sqrt{d})$. The function $|N|$ is a Euclidean function on $\mathcal{O}_K$.*

Proof. Notice that $d \equiv 1 \pmod 4$ in all of these cases. Suppose that $\alpha, \beta \in \mathcal{O}_K$ with $\beta \neq 0$. We have $\frac{\alpha}{\beta} \in K$, so we may write $\frac{\alpha}{\beta} = s + t\sqrt{d}$ for some $s, t \in \mathbb{Q}$. Fix $n \in \mathbb{Z}$ closest to $2t \in \mathbb{Q}$. We then have $|2t - n| \leq \frac{1}{2}$ and hence

$$\left|t - \frac{n}{2}\right| \leq \frac{1}{4}$$

Let $m \in \mathbb{Z}$ be the integer closest to $2s \in \mathbb{Q}$ which has the same parity as n. We then have $|2s - m| \leq 1$ and hence

$$\left|s - \frac{m}{2}\right| \leq \frac{1}{2}$$

Let $\gamma = \frac{m}{2} + \frac{n}{2}\sqrt{d} \in \mathcal{O}_K$. We then have

$$\begin{aligned}\left|N\left(\frac{\alpha}{\beta} - \gamma\right)\right| &= \left|N\left((s + t\sqrt{d}) - \left(\frac{m}{2} + \frac{n}{2}\sqrt{d}\right)\right)\right| \\ &= \left|N\left(\left(s - \frac{m}{2}\right) + \left(t - \frac{n}{2}\right)\sqrt{d}\right)\right| \\ &= \left|\left(s - \frac{m}{2}\right)^2 - d\left(t - \frac{n}{2}\right)^2\right|\end{aligned}$$

Now if $d \in \{-11, -7, -3\}$, then

$$\begin{aligned}\left|N\left(\frac{\alpha}{\beta} - \gamma\right)\right| &= |(s - \frac{m}{2})^2 - d\left(t - \frac{n}{2}\right)^2| \\ &\leq \left(s - \frac{m}{2}\right)^2 + |d| \cdot \left(t - \frac{n}{2}\right)^2 \\ &\leq \frac{1}{4} + \frac{|d|}{16} \\ &< 1\end{aligned}$$

If $d \in \{5, 13\}$, then

$$\begin{aligned}\left|N\left(\frac{\alpha}{\beta} - \gamma\right)\right| &= |(s - \frac{m}{2})^2 - d(t - \frac{n}{2})^2| \\ &\leq \max\{\left(s - \frac{m}{2}\right)^2, d\left(t - \frac{n}{2}\right)^2\} \\ &\leq \max\{\frac{1}{4}, \frac{d}{16}\} \\ &< 1\end{aligned}$$

Therefore, $|N|$ is a Euclidean function by the above lemma. □

Corollary 5.4.5. *Let $d \in \{-11, -7, -3, -2, -1, 2, 3, 5, 13\}$ and let $K = \mathbb{Q}(\sqrt{d})$. The ring $\mathcal{O}_K$ is a Euclidean domain (with Euclidean function $|N|$), hence $\mathcal{O}_K$ is a PID and a UFD.*

Proposition 5.4.6. *Let $K = \mathbb{Q}(\sqrt{10})$. The ring $\mathcal{O}_K$ is not a UFD.*

Proof. Notice that

$$2 \cdot 3 = 6 = (4 + \sqrt{10})(4 - \sqrt{10})$$

We first need to show that each of these factors are irreducible. Suppose that $\alpha, \beta \in \mathcal{O}_K$ with $2 = \alpha\beta$. We then have

$$4 = N(2) = N(\alpha\beta) = N(\alpha)N(\beta)$$

Thus, $N(\alpha), N(\beta) \in \{\pm 1, \pm 2, \pm 4\}$. If either $N(\alpha) = \pm 1$ or $N(\beta) = \pm 1$, then either α or β is a unit. Thus, we need only show that there is no $\alpha \in \mathcal{O}_K$ of norm ± 2. Suppose that $\alpha = a + b\sqrt{10}$ where $a, b \in \mathbb{Z}$ and $N(\alpha) = \pm 2$. We then have that $a^2 - 10b^2 = \pm 2$, so in particular we have $a^2 \equiv \pm 2 \pmod{10}$. Looking at the squares modulo 10 we get

$$0 \quad 1 \quad 4 \quad 9 \quad 6 \quad 5 \quad 6 \quad 9 \quad 4 \quad 1$$

Thus, there is no $a \in \mathbb{Z}$ with $a^2 \equiv \pm 2 \pmod{10}$. It follows that there is no $\alpha \in \mathcal{O}_K$ with $N(\alpha) = \pm 2$, and hence α is irreducible in $\mathcal{O}_K$. Similarly, 3 is irreducible in $\mathcal{O}_K$ because there is no $\alpha \in \mathcal{O}_K$ with $N(\alpha) = \pm 3$ (such an α would imply the existence of an $a \in \mathbb{Z}$ with $a^2 \equiv \pm 3 \pmod{10}$). Now $N(4 \pm \sqrt{10}) = 16 - 10 = 6$, so if $4 \pm \sqrt{10}$ was reducible, this would imply that there was an element of $\mathcal{O}_K$ with norm either ± 2 or ± 3, which we just showed did not exist. Therefore, each of the above factors are irreducible in $\mathcal{O}_K$. Finally, notice that neither 2 nor 3 is an associate of $4 \pm \sqrt{10}$ because if $\alpha = \beta\mu$ for a unit μ, then $N(\alpha) = N(\beta)N(\mu) = \pm N(\beta)$. □

Theorem 5.4.7. *The only integer solutions to the equation $x^3 = y^2 + 2$ are $(3, \pm 5)$.*

Proof. Suppose that (x, y) is a solution to $y^2 + 2 = x^3$. Suppose first that y is even. Working in $\mathbb{Z}$, we then have that $2 \mid (y^2 + 2)$, hence $2 \mid x^3$ and so $2 \mid x$ as 2 is prime. It follows that $8 \mid x^3$, and thus $8 \mid (y^2 + 2)$. Thus, $4 \mid (y^2 + 2)$, so as $4 \mid y^2$ (because y is even), we conclude that $4 \mid 2$, which is a contradiction.

Suppose then that y is odd. Let $R = \mathcal{O}_{\mathbb{Q}(\sqrt{-2})} = \mathbb{Z}[\sqrt{-2}]$. We know that R is a PID and UFD from above. Working in the ring R, we then have

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

We claim claim $y + \sqrt{-2}$ and $y - \sqrt{-2}$ are relatively prime in R. Suppose that $\delta \in R$ is a common divisor of $y + \sqrt{-2}$ and $y - \sqrt{-2}$. We would then have that δ divides the sum $(y + \sqrt{-2}) + (y - \sqrt{-2}) = 2y$ and the difference $(y + \sqrt{-2}) - (y - \sqrt{-2}) = 2\sqrt{-2}$. Taking norms, we conclude that

$$N(\delta) \mid 4y^2 \qquad \text{and} \qquad N(\delta) \mid 8$$

in $\mathbb{Z}$. Since y is odd, it follows that $N(\delta) \mid 4$ in $\mathbb{Z}$ and hence $N(\delta) \in \{1, 2, 4\}$ (recall that the norm of every element of R is nonnegative). A simple check shows that elements of R of norm 2 are $\pm\sqrt{-2}$ and the elements of R of norm 4 are ± 2. Notice that ± 2 does not divide $y + \sqrt{-2}$ in R (because the coefficient of $\sqrt{-2}$ is odd) and also $\pm\sqrt{-2}$ does not divide $y + \sqrt{-2}$ in R (because $\sqrt{-2}(a + b\sqrt{-2}) = -2b + a\sqrt{-2}$ and y is odd). It follows that the only common divisors of $y + \sqrt{-2}$ and $y - \sqrt{-2}$ are ± 1, so $y + \sqrt{-2}$ and $y - \sqrt{-2}$ are relatively prime in R.

Now recall that R is a UFD, so since the product $(y + \sqrt{-2})(y - \sqrt{-2})$ is a cube in R and the factors are relatively prime, it follows that there exists $\mu \in U(R)$ and $\alpha \in R$ with

$$y + \sqrt{-2} = \mu\alpha^3$$

Since $U(R) = \{\pm 1\}$, all elements of $U(R)$ are cubes in R, so we may assume that $\mu = 1$. Letting $\alpha = a+b\sqrt{-2}$ with $a, b \in \mathbb{Z}$, we have

$$\begin{aligned} y + \sqrt{-2} &= \alpha^3 \\ &= (a + b\sqrt{-2})^3 \\ &= a^3 + 3a^2b\sqrt{-2} - 6ab^2 - 2b^3\sqrt{-2} \\ &= (a^3 - 6ab^2) + (3a^2b - 2b^3)\sqrt{-2} \end{aligned}$$

Comparing the coefficients of $\sqrt{-2}$, we conclude that

$$1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2)$$

Thus, $b \mid 1$ in $\mathbb{Z}$, so $b = \pm 1$. If $b = -1$, then $1 = -(3a^2 - 2)$, so $3a^2 = 1$, a contradiction. Suppose that $b = 1$. We then have $1 = 3a^2 - 2$, so $a = \pm 1$. Thus either

$$y + \sqrt{-2} = (1 + \sqrt{-2})^3 = 1 + 3\sqrt{-2} - 6 - 2\sqrt{-2} = -5 + \sqrt{-2}$$

or

$$y + \sqrt{-2} = (-1 + \sqrt{-2})^3 = -1 + 3\sqrt{-2} + 6 - 2\sqrt{-2} = 5 + \sqrt{-2}$$

It follows that $y = \pm 5$. Hence $x^3 = 27$ and so $x = 3$. Checking the pairs $(3, \pm 5)$, we see that indeed they are solutions. □

5.5 The Eisenstein Integers

Let $K = \mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta_3)$ and let $R = \mathcal{O}_K$. We know that

$$R = \mathbb{Z}[\zeta_3] = \mathbb{Z}[\zeta_6]$$

and in this section we will view $R = \mathbb{Z}[\zeta_3]$. Now we know that $[K : \mathbb{Q}] = 2$ and that ζ_3 is a root of $x^3 - 1 = (x-1)(x^2 + x + 1)$. Thus, ζ_3 is a root of $x^2 + x + 1$. Since this polynomial has no rational roots, it is irreducible over $\mathbb{Q}$ and hence is the minimal polynomial of ζ_3 over $\mathbb{Q}$. It follows that $\{1, \zeta_3\}$ form a basis for K over $\mathbb{Q}$ and that

$$R = \{a + b\zeta_3 : a, b \in \mathbb{Z}\}$$

To see how to multiply elements of R, notice that

$$\zeta_3^2 + \zeta_3 + 1 = 0$$

from above, so

$$\zeta_3^2 = -1 - \zeta_3$$

Therefore, given $a, b, c, d \in \zeta_3$, we have

$$\begin{aligned} (a + b\zeta_3)(c + d\zeta_3) &= ac + ad\zeta_3 + bc\zeta_3 + bd\zeta_3^2 \\ &= ac + ad\zeta_3 + bc\zeta_3 + bd(-1 - \zeta_3) \\ &= (ac - bd) + (ad + bc - bd)\zeta_3 \end{aligned}$$

Also, given $a, b \in \mathbb{Z}$ we have

$$\begin{aligned} N(a+b\zeta_3) &= N\left(a+b\cdot\left(\frac{-1+\sqrt{-3}}{2}\right)\right) \\ &= N\left(\left(a-\frac{b}{2}\right)+\frac{b}{2}\cdot\sqrt{-3}\right) \\ &= \left(a-\frac{b}{2}\right)^2+3\cdot\left(\frac{b}{2}\right)^2 \\ &= a^2-ab+\frac{b^2}{4}+\frac{3b^2}{4} \\ &= a^2-ab+b^2 \end{aligned}$$

Written in these terms, since $\zeta_6 = 1+\zeta_3$, notice that

$$U(R) = \{1, -1, \zeta_3, -\zeta_3, 1+\zeta_3, -1-\zeta_3\}$$

Lemma 5.5.1. *Let $\alpha \in R$. If $N(\alpha)$ is prime in $\mathbb{Z}$, then α is irreducible and prime in R.*

Proof. Same as in the Gaussian Integers. □

Notice that $N(2) = 4$. Suppose that $2 = \alpha\beta$ where $\alpha, \beta \in R$. We then must have $N(\alpha) \in \{1, 2, 4\}$. Notice that $N(\alpha) = 2$ is impossible as follows. If $a, b \in \mathbb{Z}$, then

$$N(a+b\sqrt{-3}) = a^2+3b^2$$

and there are no solutions to $a^2+3b^2 = 2$ in $\mathbb{Z}$. Also, if $a, b \in \mathbb{Z}$ are both odd, then we still get a contradiction because

$$\frac{a^2}{4}+3\cdot\frac{b^2}{4} = 2$$

implies that $a^2+3b^2 = 8$ which has no solutions in odd integers. Thus, 2 is irreducible in R.

The case for 3 is more interesting. We clearly have

$$3 = (-1)\cdot(\sqrt{-3})^2$$

where $N(\sqrt{-3}) = 3$, so $\sqrt{-3}$ is irreducible in R. Now

$$\sqrt{-3} = 1+2\zeta_3$$

so we can write

$$3 = (-1)\cdot(1+2\zeta_3)^2$$

Another way to factor 3 is as follows. Consider the polynomial x^3-1. We know the three roots, so

$$x^3-1 = (x-1)(x-\zeta_3)(x-\zeta_3^2)$$

And hence

$$x^2+x+1 = (x-\zeta_3)(x-\zeta_3^2)$$

Plugging 1 into both sides we conclude that

$$3 = (1-\zeta_3)(1-\zeta_3^2)$$

Now we know that

$$1-\zeta_3^2 = 1-(-1-\zeta_3) = 2+\zeta_3$$

so

$$3 = (1 - \zeta_3)(2 + \zeta_3)$$

Using the fact that R is a UFD, it follows that $1 - \zeta_3$, $2 + \zeta_3$, and $1 + 2\zeta_3$ must all be associates. Recall that R has 6 units and that any associate of an element of norm 3 will also have norm 3. In fact, R has exactly 6 elements of norm 3, namely

$$-1 + \zeta_3 \quad 2 + \zeta_3 \quad -2 - \zeta_3 \quad 1 - \zeta_3 \quad 1 + 2\zeta_3 \quad -1 - 2\zeta_3$$

Since there are 6 such elements and 6 units, these all must be associates in R. Partially for historical reasons, and partially because of attempts to solve Fermat's Last Theorem (for all values of n), one typically chooses $1 - \zeta_3$ as the representative from this list. From above, we know that

$$3 = (1 - \zeta_3)(2 + \zeta_3)$$

After some work, one discovers that

$$2 + \zeta_3 = (1 + \zeta_3)(1 - \zeta_3)$$

so

$$3 = (1 + \zeta_3) \cdot (1 - \zeta_3)^2$$

where $1 + \zeta_3 \in U(R)$ and $1 - \zeta_3$ is irreducible in R. Compare this to the fact that in the Gaussian Integers $\mathbb{Z}[i]$ we have $2 = (-i) \cdot (1 + i)^2$ or $2 = i \cdot (1 - i)^2$.

Recall that we found all integers solutions to $x^2 + y^2 = z^2$ (i.e. the Pythagorean triples) by factoring the left-hand side over $\mathbb{Z}[i]$. Fermat's Last Theorem is the claim that 2 is the largest exponent n for which $x^n + y^n = z^n$ has solutions in $\mathbb{N}^+$. It is not hard to see that in order to prove this, it suffices to show that $x^4 + y^4 = z^4$ has no solutions in $\mathbb{N}^+$ and that $x^p + y^p = z^p$ has no solutions in $\mathbb{N}^+$ for odd primes p. The first of these can be handled using elementary techniques or the classification of Pythagorean triples. However, the latter is very difficult for any odd prime p. By working in the Eisenstein integers, it is possible to prove Fermat's Last Theorem for exponent 3.

Before diving into this (difficult) result, we make a few general comments about attempted proofs of Fermat's Last Theorem. Let p be an odd prime. Instead of working with the equation $x^p + y^p = z^p$, we instead work with the more symmetric equation $x^p + y^p + z^p = 0$ over $\mathbb{Z}$.

Lemma 5.5.2. *Let p be an odd prime. The equation $x^p + y^p = z^p$ has a solution in $\mathbb{N}^+$ if and only if $x^p + y^p + z^p = 0$ has a solution in $\mathbb{Z}\backslash\{0\}$.*

Proof. If $x, y, z \in \mathbb{N}^+$ satisfy $x^p + y^p = z^p$, we then have $x^p + y^p + (-z)^p = 0$ and each of $x, y, -z \in \mathbb{Z}\backslash\{0\}$ (here we are using the fact that p is odd). Conversely, suppose that $x, y, z \in \mathbb{Z}\backslash\{0\}$ satisfy $x^p + y^p + z^p = 0$. It is not possible that all of terms are positive, and it is not possible that all of the terms are negative. If one of the terms is negative and the other two are positive, we can multiply through by $(-1)^p = -1$ to find another solution with two terms negative and one term positive. Bringing the two negative terms to the other side gives a solution to $a^p + b^p = c^p$ where $a, b, c \in \mathbb{N}^+$. □

Suppose that $x, y, z \in \mathbb{Z}\backslash\{0\}$ are such that $x^p + y^p + z^p = 0$. If we are trying to rule out solutions in $\mathbb{Z}\backslash\{0\}$, it suffices to rule out such solutions where $\gcd(x, y, z) = 1$ because we can always divide through by a greatest common divisor cubed to get another solution with this property. Furthermore, notice that if $\gcd(x, y) \neq 1$ and q is a common prime divisor of x and y, then $q \mid z^p$ and hence $q \mid z$, so $\gcd(x, y, z) \neq 1$. A similar argument works for any other pair. Thus, we may assume that x, y, z are relatively prime in pairs. From here, most attempts to solve Fermat's Last Theorem break into cases. Case 1 is where p does not divide any of x, y, or z, while Case 2 is where p divides exactly one of x, y, or z (this suffices by the above comments). Case 1 is typically easier than Case 2.

We now turn our attention to the case where $p = 3$. It turns out that Case 1 can be handled by elementary means. Suppose that $x, y, z \in \mathbb{Z}\backslash\{0\}$ with $x^3 + y^3 + z^3 = 0$. Just as looking modulo 4 is a smart choice for sums of squares, it turns out that looking modulo 9 is a smart choice here.

Lemma 5.5.3. *If $a \in \mathbb{Z}$ and $3 \nmid a$, then $a^3 \equiv \pm 1 \pmod 9$.*

Proof. This can be checked by direct computation modulo 9, but here is a faster argument that we will generalize below. We have $a \equiv \pm 1 \pmod 3$. Using Lemma 2.7.2, we conclude that $a^3 \equiv (\pm 1)^3 \pmod{3^2}$, which implies that $a^3 \equiv \pm 1 \pmod 9$. □

Now it is also easily seen that if $3 \mid a$, then $a^3 \equiv 0 \pmod 9$. In other words, the cubes modulo 9 are exactly $-1, 0, 1$. Suppose now that $x, y, z \in \mathbb{Z}\backslash\{0\}$ with $x^3 + y^3 + z^3 = 0$. Suppose that none of x, y, z is divisible by 3. Working modulo 9, the previous lemma tells us that

$$\pm 1 \pm 1 \pm 1 \equiv 0 \pmod 9$$

which is a contradiction. Hence, $x^3 + y^3 + z^3 = 0$ has no Case 1 solutions.

Ruling out Case 2 solutions is much more difficult. Suppose that we have a Case 2 solution where x, y, z are relatively prime and where $3 \mid z$ (we may assume that $3 \mid z$ by symmetry and renaming of terms if necessary). We then have that

$$x^3 + y^3 = -z^3$$

The key fact is that we can factor the left-hand side in R. We have

$$\begin{aligned} x^3 + y^3 &= (x + y)(x^2 - xy + y^2) \\ &= (x + y)(x + \zeta_3 y)(x + \zeta_3^2 y) \end{aligned}$$

where we have used the fact that $\zeta_3 + \zeta_3^2 = -1$. Thus

$$(x + y)(x + \zeta_3 y)(x + \zeta_3^2 y) = -z^3$$

One might now be tempted to show that the factors on the left are relatively prime in R to argue that they are cubes. However, this doesn't work. In what follows, let $\pi = 1 - \zeta_3 \in R$ and note that π is irreducible/prime in R from above. We know that $3 \mid z$, so as $3 = (1 + \zeta_3) \cdot \pi^2$, it follows that $\pi \mid z$. Thus, π divides the right-hand side, so as π is prime, it must divide one of the factors on the left. It is not difficult to see that the three factors on the left are congruent modulo π, so since one of the factors on the left is divisible by π, all three must be. However, this suggests a potential line of attack by removing the common factor of π across all terms. Dividing both sides by π^3 we obtain

$$\frac{x + y}{\pi} \cdot \frac{x + \zeta_3 y}{\pi} \cdot \frac{x + \zeta_3^2 y}{\pi} = -\left(\frac{z}{\pi}\right)^3$$

Using the fact that x, y, z are relatively prime, one can show that the factors on the left are pairwise relatively prime in R. When multiplied together, these pairwise relatively prime elements equal a unit times a cube, so each of them must be units times cubes. From here, the key step is to notice that

$$\zeta_3^2 \cdot \left(\frac{x + y}{\pi}\right) + \left(\frac{x + \zeta_3 y}{\pi}\right) + \zeta_3 \cdot \left(\frac{x + \zeta_3^2 y}{\pi}\right) = 0$$

We now have a sum of units times cubes equal to 0 in R.

Recapping, the big picture idea is that if we have a solution, then we can divide through by π to obtain a "smaller" solution. Thus, the plan is to argue that since you can not divide by π indefinitely, there must not be a solution at all.

Now some problems worries immediately arise. Even though $x, y, z \in \mathbb{Z}$, we started to work with elements of R quite quickly and our "new solution" in R involving units.

, and then hope to do some manipulations to argue that we obtain a "smaller" solution in some sense.

In order to make this work, we have to generalize the whole apparatus to working in R throughout. Thus, we instead aim to show that

$$\alpha^3 + \beta^3 = \gamma^3$$

has no solutions in R with all three of α, β, and γ nonzero. To add some nice symmetry to the problem, we instead show that

$$\alpha^3 + \beta^3 + \gamma^3 = 0$$

has no solutions in R with all three of α, β, and γ nonzero (this suffices because $(-1)^3 = -1$ and so we can absorb the -1 into the cube). In fact, in order to prove this, we will need to prove a slightly stronger theorem involving some units (see below).

By generalizing the problem in this way, we can work in R throughout. Our first step is to prove that in such an equation, at least one of α, β, or γ is divisible by π, which is the analogue of proving one of the terms is divisible by 3 in the integer case. We first prove an important lemma (again think of π as playing a similar role to 3).

Lemma 5.5.4. *For all $\alpha \in R$, either $\alpha \equiv 1 \pmod{\pi}$, $\alpha \equiv 0 \pmod{\pi}$, or $\alpha \equiv -1 \pmod{\pi}$.*

Proof. Let $\alpha \in R$ be arbitrary. Since R is a Euclidean domain, we may fix $\gamma, \rho \in R$ with $\alpha = \gamma\pi + \rho$ and either $\rho = 0$ or $N(\rho) < N(\pi)$. Notice that $N(\pi) = 3$ and there are no elements of R with norm 2 (as discussed above), so either $\rho = 0$ or $N(\rho) = 1$. Since the elements of norm 1 are units, it follows that

$$\rho \in \{0\} \cup U(R) = \{0, 1, -1, \zeta_3, 1 + \zeta_3, -\zeta_3, -1 - \zeta_3\}$$

Now $\pi \mid (\alpha - \rho)$, so $\alpha \equiv \rho \pmod{\pi}$. Thus, it suffices to show each of the elements in the above set are congruent to one of $-1, 0, 1$ modulo π. Notice that

$$\zeta_3 \equiv 1 \pmod{\pi}$$

trivially because $\zeta_3 - 1 = -\pi$. Thus, $1 + \zeta_3 \equiv 2 \equiv -1 \pmod{\pi}$ because $\pi \mid 3$. The other two are now immediate. □

Corollary 5.5.5. *For any $\beta \in R$, $\pi \mid (\beta - \zeta_3\beta^3)$.*

Proof. Let $\beta \in R$. By the previous lemma, we know that one of $\beta \equiv 1 \pmod{\pi}$, $\beta \equiv 0 \pmod{\pi}$, or $\beta \equiv -1 \pmod{\pi}$ is true. If $\beta \equiv 1 \pmod{\pi}$, then

$$\beta - \zeta_3\beta^3 \equiv 1 - \zeta_3 \equiv 0 \pmod{\pi}$$

If $\beta \equiv 0 \pmod{\pi}$, then

$$\beta - \zeta_3\beta^3 \equiv 0 - 0 \equiv 0 \pmod{\pi}$$

If $\beta \equiv -1 \pmod{\pi}$, then

$$\beta - \zeta_3\beta^3 \equiv -1 + \zeta_3 \equiv 0 \pmod{\pi}$$

The result follows. □

Lemma 5.5.6. *Let $\alpha \in R$ with $\alpha \not\equiv 0 \pmod{\pi}$. We have the following:*

1. $\alpha^3 \equiv \pm 1 \pmod{\pi^4}$
2. $\alpha^3 \equiv \pm 1 \pmod{9}$

Proof. We have

$$3 = (1 + \zeta_3) \cdot \pi^2$$

Since $(1 + \zeta_3)^{-1} = -\zeta_3$, it follows that

$$\pi^2 = -3\zeta_3$$

Since $\alpha \not\equiv 0 \pmod{\pi}$, we know that either $\alpha \equiv 1 \pmod{\pi}$ or $\alpha \equiv -1 \pmod{\pi}$.

Suppose that $\alpha \equiv 1 \pmod{\pi}$. Fix $\beta \in R$ with $\alpha = 1 + \beta\pi$. We then have

$$\begin{aligned}\alpha^3 &= (1 + \beta\pi)^3 \\ &= 1 + 3\beta\pi + 3\beta^2\pi^2 + \beta^3\pi^3 \\ &= 1 + 3\beta^2\pi^2 + 3\beta\pi + \beta^3 \cdot (-3\zeta_3)\pi \\ &= 1 + 3\beta^2\pi^2 + 3\pi(\beta - \zeta_3\beta^3)\end{aligned}$$

Now $\pi^2 \mid 3$ and $\pi \mid (\beta - \zeta_3\beta^3)$ by the previous lemma, so the final two summands are divisible by π^4. Thus $\alpha^3 \equiv 1 \pmod{\pi^4}$.

Suppose that $\alpha \equiv -1 \pmod{\pi}$. Fix $\beta \in R$ with $\alpha = 1 + \beta\pi$. We then have

$$\begin{aligned}\alpha^3 &= (-1 + \beta\pi)^3 \\ &= -1 + 3\beta\pi - 3\beta^2\pi^2 + \beta^3\pi^3 \\ &= -1 + 3\beta^2\pi^2 - 3\beta\pi + \beta^3 \cdot (-3\zeta_3)\pi \\ &= -1 + 3\beta^2\pi^2 - 3\pi(\beta - \zeta_3\beta^3)\end{aligned}$$

Now $\pi^2 \mid 3$ and $\pi \mid (\beta - \zeta_3\beta^3)$ by the previous lemma, so the final two summands are divisible by π^4. Thus $\alpha^3 \equiv -1 \pmod{\pi^4}$.

The latter follows from the fact that $9 = (1 + \zeta_3)^2 \cdot \pi^4 = -\zeta_3 \cdot \pi^4$, so 9 and π^4 are associates. □

We now state the big theorem involving units.

Theorem 5.5.7. *There do not exist pairwise relatively prime nonzero $\alpha, \beta, \gamma \in R$ and $\mu \in U(R)$ such that $\pi \mid \gamma$ and $\alpha^3 + \beta^3 + \mu\gamma^3 = 0$.*

Before diving into the proof, we obtain the corollaries.

Corollary 5.5.8. *If $\alpha, \beta, \gamma \in R$ satisfy $\alpha^3 + \beta^3 + \gamma^3 = 0$, then at least one of α, β, γ equals 0.*

Proof. Suppose that there exist $\alpha, \beta, \gamma \in R\backslash\{0\}$ with $\alpha^3 + \beta^3 + \gamma^3 = 0$. Fix a greatest common divisor δ of $\{\alpha, \beta, \gamma\}$ (recall these exists in PIDs even for multiple elements). Note that $\delta \neq 0$ because α, β, and γ are nonzero. Fix $\alpha', \beta', \gamma' \in R$ with $\alpha = \delta\alpha'$, $\beta = \delta\beta'$, and $\gamma = \delta = \gamma'$. We then have

$$(\alpha')^3 + (\beta')^3 + (\gamma')^3 = 0$$

A simple check shows that the only common divisors of α', β', and γ' are units. Now if any two of these have a common prime divisor, then it must divide the other, so in fact α', β', and γ' are pairwise relatively prime.

We now argue that at least one of α', β', or γ' is divisible by π. Suppose not. Reducing the equation modulo 9 in R, we then have

$$\pm 1 \pm 1 \pm 1 \equiv 0 \pmod{9}$$

Since $9 \nmid \pm 1$ and $9 \nmid \pm 3$ in R, this is a contradiction. Therefore, at least of α', β', or γ' is divisible by π. This contradicts the previous theorem with $\mu = 1$ (changing the role of the terms if necessary). □

Corollary 5.5.9. *If $x, y, z \in \mathbb{Z}$ satisfy $x^3 + y^3 = z^3$, then at least one of x, y, z equals 0.*

Proof. We have $x^3 + y^3 + (-z)^3$, so we can apply the previous corollary. □

We now jump into a proof of the theorem.

Proof of 5.5.7. exist pairwise relatively prime nonzero $\alpha, \beta, \gamma \in R$ and $\mu \in U(R)$ such that $\pi \mid \gamma$ and $\alpha^3 + \beta^3 + \mu\gamma^3 = 0$. Pick such a choice one with $ord_\pi \gamma$ as small as possible, say $ord_\pi \gamma = n$ where $n \in \mathbb{N}^+$. Notice that we must have $ord_\pi \alpha = 0$ and $ord_\pi \beta = 0$ since we are assuming that the terms are relatively prime.

We first claim that $n \geq 2$. Suppose instead that $n = 1$ and write $\gamma = \pi \cdot \gamma'$ where $\pi \nmid \gamma'$.

$$\alpha^3 + \beta^3 + \pi^3 \cdot (\gamma')^3 = 0$$

Reducing this equation modulo 9, we conclude that

$$\pm 1 \pm 1 \pm \pi^3 \equiv 0 \pmod 9$$

Since

$$\pi^3 = -3 - 6\zeta$$

it follows that

$$\pm 1 \pm 1 \pm (3 + 6\zeta) \equiv 0 \pmod 9$$

A simple check of norms shows that this is impossible. Therefore, $n \geq 2$.

Now we have $\alpha^3 + \beta^3 + \gamma^3 = 0$ and $\pi \mid \gamma$. It follows that $\pi \mid (\alpha^3 + \beta^3)$. Since

$$\alpha^3 + \beta^3 = (\alpha + \beta)(\alpha + \zeta_3\beta)(\alpha + \zeta_3^2\beta)$$

and π is prime, we conclude that π divides one of the terms on the right. However, notice that

$$(\alpha + \beta) - (\alpha + \zeta_3\beta) = (1 - \zeta_3)\beta = \pi\beta$$

so

$$\alpha + \beta \equiv \alpha + \zeta_3\beta \pmod \pi$$

We also have

$$(\alpha + \zeta_3\beta) - (\alpha + \zeta_3^2\beta) = \zeta_3\beta - \zeta_3^2\beta = \pi\zeta_3\beta$$

so

$$(\alpha + \zeta_3\beta) \equiv (\alpha + \zeta_3^2\beta) \pmod \pi$$

Combining the last two, we conclude that

$$\alpha + \beta \equiv \alpha + \zeta_3^2\beta \pmod \pi$$

Thus, all three of the factors are congruent modulo π. Since π divides at least one of them it follows that π divides all of them.

Now we claim that the elements

$$\frac{\alpha + \beta}{\pi}, \frac{\alpha + \zeta_3\beta}{\pi}, \frac{\alpha + \zeta_3^2\beta}{\pi}$$

are pairwise relatively prime. Suppose that δ is a common divisor or the first two. We then have that δ divides their difference, which is $\frac{(1-\zeta_3)\beta}{\pi} = \beta$. We also have that δ divides the second minus ζ_3 times the first, which is $\frac{(1-\zeta_3)\alpha}{\pi} = \alpha$. Since α and β are relatively prime, it follows that δ is a unit. The other two pairs are treated similarly.

Dividing both sides of the equation by π^3, we obtain

$$\frac{\alpha+\beta}{\pi}\cdot\frac{\alpha+\zeta_3\beta}{\pi}\cdot\frac{\alpha+\zeta_3^2\beta}{\pi}=(-\mu)\cdot\left(\frac{\gamma}{\pi}\right)^3$$

where each of the factors on the left are pairwise relatively prime. Since the product on the left is a unit times a cube, and each of the factors on the left are pairwise relatively prime, it follows that each of these factors are units times cubes (here we are using that R is a UFD). Thus we can write

$$\frac{\alpha+\beta}{\pi}=\varepsilon_1\rho^3 \qquad \frac{\alpha+\zeta_3\beta}{\pi}=\varepsilon_2\sigma^3 \qquad \frac{\alpha+\zeta_3^2\beta}{\pi}=\varepsilon_3\tau^3$$

where the $\varepsilon_i\in R$ are units. Notice that ρ, σ, and τ are pairwise relatively prime. Since $n\geq 2$, the right-hand side is still divisible by π, so at least of σ, ρ, τ is divisible π. By relabeling, we can assuming that $\pi\mid\tau$. Notice that $\pi\nmid\rho$ and $\pi\nmid\sigma$ because they are relatively prime. Thus, we must have $ord_\pi\tau=n-1$.

Now

$$\zeta_3^2\cdot\frac{\alpha+\beta}{\pi}+\frac{\alpha+\zeta_3\beta}{\pi}+\zeta_3\cdot\frac{\alpha+\zeta_3^2\beta}{\pi}=0$$

so

$$\zeta_3^2\varepsilon_1\rho^3+\varepsilon_2\sigma^3+\zeta_3\varepsilon_3\tau^3=0$$

Dividing through by the unit $\zeta_3^2\varepsilon_1$, we obtain

$$\rho^3+\varepsilon_4\sigma^3+\varepsilon_5\tau^3=0$$

To finish the argument in contradiction, it suffices to show that $\varepsilon_4=\pm 1$. Reducing modulo π^3, we obtain

$$\pm 1\pm\varepsilon_4\equiv 0 \pmod{\pi^3}$$

A simple check of units shows that this is only possible if $\varepsilon_4=\pm 1$. Absorbing the negative, we obtain

$$\rho^3+\sigma^3+\varepsilon_5\tau^3=0$$

where $ord_\pi\tau=n-1$, a contradiction. □

Chapter 6

Quadratic Reciprocity

6.1 Quadratic Residues and the Legendre Symbol

Definition 6.1.1. *Let $n \in \mathbb{N}^+$ and let $a \in \mathbb{Z}$ with $\gcd(a,n) = 1$. We say that a is a* quadratic residue *modulo n if there exists $x \in \mathbb{Z}$ with $x^2 \equiv a \pmod{n}$ (equivalently, if $\overline{a}$ is a square in $U(\mathbb{Z}/n\mathbb{Z})$).*

Definition 6.1.2. *Suppose that $p \in \mathbb{N}^+$ is prime. Define $Q_p \subseteq U(\mathbb{Z}/p\mathbb{Z})$ to be*

$$Q_p = \{\overline{a} : a \text{ is a quadratic residue modulo } p\}$$

so Q_p is the set of cosets of quadratic residues.

We have already studied quadratic residues if you don't remember doing so. In Homework 4, Problem 6, and Homework 5, Problem 1, you proved the following.

Theorem 6.1.3. *Suppose that $p, m \in \mathbb{N}^+$ where p is prime. Let $d = \gcd(m, p-1)$. Define $\psi\colon U(\mathbb{Z}/p\mathbb{Z}) \to U(\mathbb{Z}/p\mathbb{Z})$ by letting $\psi(x) = x^m$.*

1. *ψ is a group homomorphism.*
2. *$|\ker(\psi)| = d$.*
3. *$|\text{range}(\psi)| = \frac{p-1}{d}$.*
4. *Given $\overline{a} \in U(\mathbb{Z}/p\mathbb{Z})$, we have $\overline{a} \in \text{range}(\psi)$ if and only if $a^{(p-1)/d} \equiv 1 \pmod{p}$.*

If p is an odd prime and $m = 2$, we have $d = \gcd(2, p-1) = 2$, so we obtain the following fundamental corollary (which we reprove here).

Theorem 6.1.4. *Suppose that $p \in \mathbb{N}^+$ is an odd prime. Define $\psi\colon U(\mathbb{Z}/p\mathbb{Z}) \to U(\mathbb{Z}/p\mathbb{Z})$ by letting $\psi(x) = x^2$.*

1. *ψ is a group homomorphism.*
2. *$|\ker(\psi)| = 2$ (so $\ker(\psi) = \{\overline{1}, \overline{-1}\}$).*
3. *$|\text{range}(\psi)| = \frac{p-1}{2}$.*
4. *Given $\overline{a} \in U(\mathbb{Z}/p\mathbb{Z})$, we have $\overline{a} \in \text{range}(\psi)$ if and only if $a^{(p-1)/2} \equiv 1 \pmod{p}$.*

We now pull apart this theorem into two important corollaries.

Proposition 6.1.5. *Suppose that $p \in \mathbb{N}^+$ is an odd prime. We then have that Q_p is a (multiplicative) subgroup of $U(\mathbb{Z}/p\mathbb{Z})$ and that $|Q_p| = \frac{p-1}{2}$.*

Proof. We have $Q_p = \text{range}(\psi)$, so Q_p is a subgroup because ψ is a group homomorphism. The second result is immediate. □

One can also prove the previous proposition directly without working through the general theory. One can show that Q_p is a subgroup by directly arguing that the product of two squares is a square, and that the multiplicative inverse of a square is also a square (and clearly $\overline{1}$ is a square). The second part can also be argued directly as follows. Since $a^2 \equiv (-a)^2 \pmod{p}$, we have $a^2 \equiv (p-a)^2 \pmod{p}$, so determine the squares modulo p it suffices to consider the following squares:

$$1^2 \quad 2^2 \quad 3^2 \quad \ldots \quad (\tfrac{p-1}{2})^2$$

Also, if $a^2 \equiv b^2 \pmod{p}$, then $p \mid (a^2 - b^2)$, so $p \mid (a-b)(a+b)$, hence either $p \mid (a-b)$ or $p \mid (a+b)$ (since p is prime), and thus either $a \equiv b \pmod{p}$ or $a \equiv -b \pmod{p}$. Thus, the above $\frac{p-1}{2}$ squares and pairwise not congruent modulo p.

Restating part 4 of the above theorem, we obtain the following fundamental characterization of quadratic residues modulo p.

Corollary 6.1.6 (Euler's Criterion)**.** *Let p be an odd prime. A number $a \in \mathbb{Z}$ with $p \nmid a$ is a quadratic residue modulo p if and only if*

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

Proof. This is immediate from the above theorem, but we recall the proof.

Let $a \in \mathbb{Z}$ with $p \nmid a$. Suppose first a is a quadratic residue modulo p. Fix $b \in \mathbb{Z}$ with $p \nmid b$ and $b^2 \equiv a \pmod{p}$. We then have

$$\begin{aligned} a^{(p-1)/2} &\equiv (b^2)^{(p-1)/2} \pmod{p} \\ &\equiv b^{p-1} \pmod{p} \\ &\equiv 1 \pmod{p} \end{aligned}$$

where the last line follows from Fermat's Little Theorem.

We now prove the converse by showing if a is not a quadratic residue modulo p, then $a^{(p-1)/2} \not\equiv 1 \pmod{p}$. This proof is a clever counting argument using roots of polynomials. By the direction we just proved, if a is a quadratic residue modulo p, then $\overline{a}$ is a root of the polynomial $x^{(p-1)/2} - \overline{1} \in \mathbb{Z}/p\mathbb{Z}[x]$. Since $\mathbb{Z}/p\mathbb{Z}$ is a field, the polynomial $x^{(p-1)/2} - 1$ has at most $\frac{p-1}{2}$ many roots in $\mathbb{Z}/p\mathbb{Z}$. Since $|Q_p| = \frac{p-1}{2}$ from above, we conclude that these must be all of the roots of $x^{(p-1)/2} - \overline{1}$ in $\mathbb{Z}/p\mathbb{Z}$. Therefore, if $\overline{a} \notin Q_p$, then $\overline{a}$ is not a root of this polynomial, and hence $a^{(p-1)/2} \not\equiv 1 \pmod{p}$. □

Euler's Criterion establishes an important characterization of when a an $a \in \mathbb{Z}$ is a quadratic residue modulo an odd prime p. Since raising a number to a power is computationally quite fast (using repeated squaring), this gives a satisfactory computational solution that is significantly more efficient that simply trying all of the squares. However, there is much more that can be said, including important structural connections about when a number a is quadratic residue modulo various primes.

Before jumping into these important results, we first say a bit about quadratic residues modulo n where n is not a prime. Suppose that $n \in \mathbb{N}^+$ has prime factorization

$$n = p_1^{k_1} p_2^{k_2} \cdots p_\ell^{k_\ell}$$

Let $a \in \mathbb{Z}$ with $\gcd(a, n) = 1$. Using the Chinese Remainder Theorem, it is straightforward to see that a is a quadratic residue modulo n if and only a is a quadratic residue modulo each $p_i^{k_i}$. On the homework,

you will show that for an odd prime p, we have that a is a quadratic residue modulo p^k if and only if a is a quadratic residue modulo p. It is also possible to determine the quadratic residues modulo powers of 2 using the structure of $U(\mathbb{Z}/2^k\mathbb{Z})$ (see later homework... probably). Thus, we have reduced the problem to one of determining if a is a quadratic residue modulo odd primes.

Definition 6.1.7. *Let $p \in \mathbb{N}^+$ be an odd prime. For any $a \in \mathbb{Z}$ we define*

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue modulo } p \\ -1 & \text{if } a \text{ is a quadratic nonresidue modulo } p \\ 0 & \text{if } p \mid a \end{cases}$$

This notation is called the Legendre symbol.

Proposition 6.1.8. *Suppose that p is an odd prime and $a \in \mathbb{Z}$. We have*

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

Proof. If $p \mid a$, then the left-hand side is 0 and the right-hand side is divisible by p, so the congruence is satisfied. Suppose then that $p \nmid a$. If a is quadratic residue modulo p, then $(\frac{a}{p}) = 1$ by definition and $a^{(p-1)/2} \equiv 1 \pmod{p}$ by Euler's Criterion. Suppose then that a is not a quadratic residue modulo p. By definition, we have $(\frac{a}{p}) = -1$. Now

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

by Fermat's Little Theorem. Using Proposition 2.5.2, it follows that either

$$a^{(p-1)/2} \equiv 1 \pmod{p} \qquad \text{or} \qquad a^{(p-1)/2} \equiv -1 \pmod{p}$$

Since a is not a quadratic residue modulo p, Euler's Criterion tells us that $a^{(p-1)/2} \not\equiv 1 \pmod{p}$. Thus, we must have $a^{(p-1)/2} \equiv -1 \pmod{p}$. □

We now have another proof of an old fact.

Corollary 6.1.9. *If $p \in \mathbb{N}^+$ is an odd prime, then*

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

Thus, -1 is a quadratic residue modulo p exactly when $p \equiv 1 \pmod{4}$.

Proof. We have $p \nmid -1$ because p is prime. By the previous Corollary, we know that

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

Since both the sides of this equation are ± 1 and $1 \not\equiv -1 \pmod{p}$ (because p is an odd prime), it follows that

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

To finish the argument notice that if $p \equiv 1 \pmod{4}$, then $\frac{p-1}{2}$ is even so $(-1)^{(p-1)/2} = 1$, while if $p \equiv 3 \pmod{4}$, then $\frac{p-1}{2}$ is odd so $(-1)^{(p-1)/2} = -1$ □

Corollary 6.1.10. *There are infinitely many primes $p \in \mathbb{N}^+$ with $p \equiv 1 \pmod 4$.*

Proof. Notice that at least one such prime exists, namely 5. Suppose now that $p_1, p_2, \ldots, p_k$ are primes with $p_i \equiv 1 \pmod 4$. Consider the number

$$n = (2p_1p_2 \cdots p_k)^2 + 1$$

Notice that $n \geq 2$, so n has a prime divisor. Fix a prime $q \in \mathbb{N}^+$ such that $q \mid n$. Notice that q is odd because m is odd. We then have that -1 is a quadratic residue modulo q, so by the previous corollary we know that $q \equiv 1 \pmod 4$. Now if $q = p_i$, then q would divide

$$n - (2p_1p_2 \cdots p_k)^2 = 1$$

a contradiction. Thus, we have established the existence of a prime $q \equiv 1 \pmod 4$ such that $q \neq p_i$ for all i. The result follows. □

Proposition 6.1.11. *Let p be an odd prime.*

1. *For any $a, b \in \mathbb{Z}$, we have*
$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$$
2. *If $a \equiv b \pmod p$, then*
$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$$
3. *For any a with $p \nmid a$, we have*
$$\left(\frac{a^2}{p}\right) = 1$$
4. *We always have*
$$\left(\frac{1}{p}\right) = 1$$

Proof. Properties 2, 3, and 4 follow directly from the definition of the Legendre symbol. We now prove property 1. Suppose that $a, b \in \mathbb{Z}$. If either $p \mid a$ or $p \mid b$, then both sides are 0. Suppose then that $p \nmid a$ and $p \nmid b$. We then have that

$$\begin{aligned}\left(\frac{ab}{p}\right) &\equiv (ab)^{(p-1)/2} \pmod p \\ &\equiv a^{(p-1)/2} \cdot b^{(p-1)/2} \pmod p \\ &\equiv \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \pmod p\end{aligned}$$

Therefore, the two values are equivalent modulo p. Since they both equal ± 1, and $1 \not\equiv -1 \pmod p$, they must be equal. □

The first statement in the above proposition is saying that the function $\phi\colon U(\mathbb{Z}/p\mathbb{Z}) \to \{1, -1\}$ defined by

$$\phi(a) = \left(\frac{a}{p}\right)$$

is a group homomorphism from $U(\mathbb{Z}/p\mathbb{Z})$ to $\{\pm 1\}$ (viewed as a group under multiplication). It says that the product of two quadratic residues is a quadratic residue, that the product of a quadratic residue and

a quadratic nonresidue is a quadratic nonresidue, and that the product of two quadratic nonresidues is a residue. The first two of these are not very surprising and can be easily generalized to any field. That is, if F is a field and we are working in the group $U(F) = F\backslash\{0\}$, then the product of two squares is alway a square, and the product of a square and a nonsquare is always a nonsquare (prove it!). However, it is certainly not true in general fields that the product of two nonsquares is a square, so this is something very special about $U(\mathbb{Z}/p\mathbb{Z})$.

For example, consider the field $\mathbb{Q}$ and think about the squares in $U(\mathbb{Q})$. Notice that neither 2 nor 3 are squares in $U(\mathbb{Q})$, but their product $2 \cdot 3 = 6$ is also not a square in $\mathbb{Q}$. In contrast, the product of two nonsquares in $U(\mathbb{R})$ must be a square because an element is a square exactly when it is positive.

Let's analyze this situation more generally. Let F be a field and consider the function $\phi\colon U(F) \to \{1, -1\}$ defined by

$$\phi(a) = \begin{cases} 1 & \text{if } a \text{ is a square in } U(F) \\ -1 & \text{if } a \text{ is not a square in } U(F) \end{cases}$$

As we saw above, it is not generally true that ϕ is a group homomorphism (since if $F = \mathbb{Q}$, then $h(6) = -1$ while $h(2) \cdot h(3) = (-1) \cdot (-1) = 1$). Let $A = \{a \in U(F) : \phi(a) = 1\}$ be the set of squares. It is still true in general that H is a multiplicative subgroup of $U(F)$ (check it!), and thus it must be a normal subgroup because $U(F)$ is abelian. Thus, there is a natural projection map

$$\pi\colon U(F) \to U(F)/H$$

In the case where $F = \mathbb{Z}/p\mathbb{Z}$, know that $H = Q_p$ has $\frac{p-1}{2}$ elements, so H has index 2 in $U(F)$ and hence the group $U(F)/H$ has two elements. Thus, in this case, the quotient group $U(\mathbb{Z}/p\mathbb{Z})/Q_p$ has order 2 and can be identified isomorphically with $\{\pm 1\}$ under multiplication. In this way, the map ϕ above given by $a \mapsto (\frac{a}{p})$ is the same map as π. Fundamentally, the fact that Q_p has index 2 in $U(\mathbb{Z}/p\mathbb{Z})$ is the reason why the product of two nonsquares must be a square (because any nonsquare gets mapped to -1 in this quotient, and hence the product of two nonsquares corresponds to $(-1) \cdot (-1) = 1$).

If you are working in a field F where the set of squares H has index greater than 2 in $U(F)$, you still get this projection map $\pi\colon U(F) \to U(F)/H$ that is a homomorphism, but since the quotient object has more than 2 elements there is no reason why the product of two nonidentity elements in the quotient will be the identity element, so there is no reason to believe that the product of two nonsquares will be a square.

6.2 When 2 is a Quadratic Residue Modulo p

Let p be an odd prime. We have Euler's Criterion, which tells us that 2 is a quadratic residue modulo p if and only if

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

For a large prime p, a computer can quickly determine this (by repeated squaring), but it is still a significant computation. We want to determine a simple characterization for when 2 is a quadratic residue modulo p that would allow us to immediately conclude the answer using hardly any work.

Although there are elementary ways to get at such a characterization, we give a slick proof using some algebraic number theory that will pave the way for later results. The very clever idea is to work in a ring extending the integers for which we can easily manipulate Euler's Criterion. In the integers, the value of $2^{(p-1)/2}$ is difficult to compute precisely. So the first step is to thing of a ring that has elements where large powers are easy to compute. We know that roots of unity cycle around when you take powers, so we would like to work in a ring where we express 2 in terms of roots of unity. Of course, another problem is that the power of a sum is not the sum of the powers. However, since we are working modulo a prime, this is easily overcome.

Lemma 6.2.1. *Let R be a subring of $\mathbb{C}$ and let $p \in \mathbb{N}^+$ be prime. Working in R, we have*

$$(\alpha + \beta)^p \equiv \alpha^p + \beta^p \pmod{p}$$

for all $\alpha, \beta \in R$.

Proof. By the Binomial Theorem, we have

$$(\alpha+\beta)^p = \alpha^p + \binom{p}{1}\alpha^{p-1}\beta + \binom{p}{2}\alpha^{p-2}\beta^2 + \cdots + \binom{p}{p-1}\alpha\beta^{p-1} + \beta^p$$

hence

$$(\alpha+\beta)^p - (\alpha^p+\beta^p) = \binom{p}{1}\alpha^{p-1}\beta + \binom{p}{2}\alpha^{p-2}\beta^2 + \cdots + \binom{p}{p-1}\alpha\beta^{p-1}$$

We know $p \mid \binom{p}{k}$ in $\mathbb{Z}$, and hence $p \mid \binom{p}{k}$ in R, whenever $1 \leq k \leq p-1$. Thus, the right-hand side is divisible by p in R, and therefore

$$(\alpha + \beta)^p \equiv \alpha^p + \beta^p \pmod{p}$$

in R. □

Lemma 6.2.2. *Let R be a subring of $\mathbb{C}$ consisting entirely of algebraic integers (for example, $R \subseteq \mathcal{O}_K$ for some number field K). Let $m, n \in \mathbb{Z}$. If $m \mid n$ in R, then $m \mid n$ in $\mathbb{Z}$.*

Proof. Suppose that $m \mid n$ in R. Fix $\alpha \in R$ with $n = m\alpha$. Notice that if $m = 0$, then $n = 0$, so trivially $m \mid n$ in $\mathbb{Z}$. Suppose then that $m \neq 0$. We then have that $\alpha = \frac{n}{m} \in \mathbb{Q}$. Since $\alpha \in R$, we know that α is an algebraic integer, so using Corollary 4.5.10 we can conclude that $\alpha \in \mathbb{Z}$. Therefore, $m \mid n$ in $\mathbb{Z}$. □

Corollary 6.2.3. *Let R be a subring of $\mathbb{C}$ consisting entirely of algebraic integers. Let $a, b, c \in \mathbb{Z}$. We have $a \equiv b \pmod{c}$ in $\mathbb{Z}$ if and only if $a \equiv b \pmod{c}$ in R.*

Proof. Immediate from the previous lemma. □

Now Euler's Criterion deals with $2^{(p-1)/2}$ modulo p, so in order to use these results, we instead seek to express $\sqrt{2}$ in terms of roots of unity because $2^{(p-1)/2} = (\sqrt{2})^{p-1}$, which is much closer to a p^{th} power. In a perfect world, we would ideally want to be able to express $\sqrt{2}$ as sums and differences of roots of unity without any coefficients so that raising to a power will be incredibly easy.

Thus, our goal is to express $\sqrt{2}$ in terms of roots of unity, and after playing around the ring that works out is $R = \mathbb{Z}[\zeta_8]$ where

$$\begin{aligned}
\zeta_8 &= e^{2\pi i/8} \\
&= e^{\pi i/4} \\
&= \cos\frac{\pi}{4} + i \cdot \sin\frac{\pi}{4} \\
&= \frac{\sqrt{2}}{2} + i \cdot \frac{\sqrt{2}}{2}
\end{aligned}$$

We have

$$\begin{aligned}
\zeta_8^{-1} &= \zeta_8^7 \\
&= e^{14\pi i/8} \\
&= e^{7\pi i/4} \\
&= \cos\frac{7\pi}{4} + i \cdot \sin\frac{7\pi}{4} \\
&= \frac{\sqrt{2}}{2} - i \cdot \frac{\sqrt{2}}{2}
\end{aligned}$$

hence

$$\zeta_8 + \zeta_8^7 = \zeta_8 + \zeta_8^{-1} = \sqrt{2}$$

Thus, we will work in the ring $R = \mathbb{Z}[\zeta_8]$.

Although not essential for our development, we first establish some facts about this ring. We know that ζ_8 is a root of the polynomial $x^8 - 1$. We have

$$x^8 - 1 = (x^4 - 1)(x^4 + 1)$$

Now $\zeta_8^4 = e^{\pi i} = -1$, so ζ_8 is not a root of $x^4 - 1$. It follows that ζ_8 is a root of $x^4 + 1$, and we claim that this is the minimal polynomial of ζ_8 over $\mathbb{Q}$. The polynomial $x^4 + 1$ also has no rational roots (because $(\pm 1)^4 + 1 = 2$), but we need to check that it is not the product of two degree 2 polynomials in $\mathbb{Z}[x]$. Suppose then that $a, b, c, d \in \mathbb{Z}$ with

$$\begin{aligned} x^4 + 1 &= (x^2 + ax + b)(x^2 + cx + d) \\ &= x^4 + (a + c)x^3 + (ac + b + d)x^2 + (ad + bc)x + bd \end{aligned}$$

Looking at constant terms, we then have $bd = 1$, so either $b = 1 = d$ or $b = -1 = d$. This gives two cases.

- *Case 1:* Suppose that $b = 1 = d$. Looking at the coefficients of x^3 and x^2, we have $a + c = 0$ and $ac + 2 = 0$. Thus $c = -a$ and this implies$-a^2 + 2 = 0$. From this we conclude that $a^2 = 2$ which is a contradiction because $a \in \mathbb{Z}$.
- *Case 2:* Suppose that $b = -1 = d$. Looking at the coefficients of x^3 and x^2, we have $a + c = 0$ and $ac - 2 = 0$. Thus $c = -a$ and this implies$-a^2 - 2 = 0$. From this we conclude that $a^2 = -2$ which is a contradiction because $a \in \mathbb{Z}$.

Therefore, $x^4 + 1$ is also not the product of two quadratics, and hence $x^4 + 1$ is irreducible in $\mathbb{Q}[x]$. Therefore, $x^4 + 1$ is the minimal polynomial of ζ_8 over $\mathbb{Q}$. It follows that ζ_8 is an algebraic integer and that

$$\mathbb{Q}(\zeta_8) = \{a_0 + a_1\zeta_8 + a_2\zeta_8^2 + a_3\zeta_8^3 : a_i \in \mathbb{Q}\}$$

Intuitively, we can reduce powers of ζ_8 beyond 3 by using the relation that $\zeta_8^4 = -1$. We also conclude that

$$\mathbb{Z}[\zeta_8] = \{a_0 + a_1\zeta_8 + a_2\zeta_8^2 + a_3\zeta_8^3 : a_i \in \mathbb{Z}\}$$

Since ζ_8 is an algebraic integer, we certainly have $\mathbb{Z}[\zeta_8] \subseteq \mathcal{O}_{\mathbb{Q}(\zeta_8)}$. It is not clear at this point whether equality holds (it does), but we do not need that here. We know that $\zeta_8^7 \in \mathbb{Z}[\zeta_8]$, and in terms of the above representations notice that

$$\zeta_8^7 = \zeta_8^4 \cdot \zeta_8^3 = (-1) \cdot \zeta_8^3$$

Let $\tau = \zeta_8 + \zeta_8^{-1} = \sqrt{2} \in R$. Working in R, we know that

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

so as $\tau^2 = 2$, it follows that

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

Multiplying both sides by τ we conclude that

$$\tau^p \equiv \left(\frac{2}{p}\right) \cdot \tau \pmod{p}$$

Now

$$\begin{aligned}\tau^p &= (\zeta_8 + \zeta_8^{-1})^p \\ &\equiv \zeta_8^p + \zeta_8^{-p} \pmod{p}\end{aligned}$$

We now have a few cases.

- If $p \equiv 1 \pmod 8$, then

$$\begin{aligned}\tau^p &\equiv \zeta_8^p + \zeta_8^{-p} \pmod{p} \\ &\equiv \zeta_8 + \zeta_8^{-1} \pmod{p} \\ &\equiv \tau \pmod{p}\end{aligned}$$

- If $p \equiv 3 \pmod 8$, then

$$\begin{aligned}\tau^p &\equiv \zeta_8^p + \zeta_8^{-p} \pmod{p} \\ &\equiv \zeta_8^3 + \zeta_8^{-3} \pmod{p} \\ &\equiv -\zeta_8^{-1} - \zeta_8 \pmod{p} \\ &\equiv -\tau \pmod{p}\end{aligned}$$

- If $p \equiv 5 \pmod 8$, then

$$\begin{aligned}\tau^p &\equiv \zeta_8^p + \zeta_8^{-p} \pmod{p} \\ &\equiv \zeta_8^5 + \zeta_8^{-5} \pmod{p} \\ &\equiv -\zeta_8 - \zeta_8^{-1} \pmod{p} \\ &\equiv -\tau \pmod{p}\end{aligned}$$

- If $p \equiv 7 \pmod 8$, then

$$\begin{aligned}\tau^p &\equiv \zeta_8^p + \zeta_8^{-p} \pmod{p} \\ &\equiv \zeta_8^7 + \zeta_8^{-7} \pmod{p} \\ &\equiv \zeta_8^{-1} + \zeta_8 \pmod{p} \\ &\equiv \tau \pmod{p}\end{aligned}$$

Suppose then that either $p \equiv 1 \pmod 8$ or $p \equiv 7 \pmod 8$. We then have $\tau^p \equiv \tau \pmod p$, so

$$\tau \equiv \left(\frac{2}{p}\right) \cdot \tau \pmod{p}$$

Multiplying both sides by τ and using the fact that $\tau^2 = 2$, we conclude that

$$2 \equiv \left(\frac{2}{p}\right) 2 \pmod{p}$$

So far we have been working in R, but from the above lemma this congruence holds in $\mathbb{Z}$ as well. Since $\gcd(2, p) = 1$, we may cancel the 2's from both sides to conclude that

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

and hence $(\frac{2}{p}) = 1$.

Suppose instead that either $p \equiv 3 \pmod 8$ or $p \equiv 5 \pmod 8$. We then have $\tau^p \equiv -\tau \pmod p$, so Thus

$$-\tau \equiv \left(\frac{2}{p}\right) \cdot \tau \pmod p$$

Multiplying both sides by τ and using the fact that $\tau^2 = 2$, we conclude that

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

As above, this congruence holds in $\mathbb{Z}$ as well. Since $\gcd(2, p) = 1$, we may cancel the 2's from both sides to conclude that

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

and hence $(\frac{2}{p}) = -1$.

Theorem 6.2.4. *If $p \in \mathbb{N}^+$ is an odd prime, then*

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 7 \pmod 8 \\ -1 & \text{if } p \equiv 3, 5 \pmod 8 \end{cases}$$

Stated more succinctly, we have

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$$

Proof. The only thing left to prove is the last step. If $p \equiv 1, 7 \pmod 8$, then $p^2 \equiv 1^2, 7^2 \pmod{16}$, so $p^2 \equiv 1 \pmod{16}$ and hence $\frac{p^2-1}{8}$ is even. If $p \equiv 3, 5 \pmod 8$, then $p^2 \equiv 3^2, 5^2 \pmod{16}$, so $p^2 \equiv 9 \pmod{16}$ and hence $\frac{p^2-1}{8}$ is odd. □

Corollary 6.2.5. *If $p \in \mathbb{N}^+$ is an odd prime, then*

$$\left(\frac{-2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 3 \pmod 8 \\ -1 & \text{if } p \equiv 5, 7 \pmod 8 \end{cases}$$

Proof. We know that

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

We know that $\left(\frac{-1}{p}\right) = 1$ if and only if $p \equiv 1 \pmod 4$, which is equivalent to saying that either $p \equiv 1 \pmod 8$ or $p \equiv 5 \pmod 8$. We know consider the four possible cases to obtain the result. □

Corollary 6.2.6. *Let $p \in \mathbb{N}^+$ be prime. There exists $a, b \in \mathbb{Z}$ with $p = a^2 + 2b^2$ in exactly the following cases.*

- $p = 2$
- $p \equiv 1 \pmod 8$
- $p \equiv 3 \pmod 8$

Proof. This follows from the homework using the fact that $\mathbb{Z}[\sqrt{-2}]$ is a UFD along with the previous corollary. □

6.3 Quadratic Reciprocity

We now prove the following major theorem about quadratic residues.

Theorem 6.3.1 (Quadratic Reciprocity)**.** *Let p and q be distinct odd primes. We have*

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

In other words

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

We build up to the proof in stages. Suppose that $p \in \mathbb{N}^+$. The key idea to the proof is to use the important insights of the last section and try to find sums of roots of unity that equal $\sqrt{\pm p}$. For small values of p, one can use some trigonometry to find ad hoc solutions. For example, we have

$$\zeta_3 - \zeta_3^2 = \sqrt{-3}$$

When $n = 5$, notice that

$$\zeta_5 = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}$$

and using some elementary trigonometry one can show that

$$\cos\frac{2\pi}{5} = \frac{\sqrt{5}-1}{4}$$

From here, some calculations give the following:

$$\zeta_5 - \zeta_5^2 - \zeta_5^3 + \zeta_5^4 = \sqrt{5}$$

Before jumping into the big theory, we first think about certain simple sums of roots of unity.

Lemma 6.3.2. *Let $p \in \mathbb{N}^+$ be an odd prime. We have $\sum_{k=0}^{p-1} \zeta_p^k = 0$.*

Proof. One proof is to use the sum of a finite geometric sequences. Since $\zeta_p \neq 1$, we have

$$\begin{aligned} \sum_{k=0}^{p-1} \zeta_p^k &= 1 + \zeta_p + \zeta_p^2 + \cdots + \zeta_p^{p-1} \\ &= \frac{\zeta_p^p - 1}{\zeta_p - 1} \\ &= \frac{1-1}{\zeta_p - 1} \\ &= 0 \end{aligned}$$

Alternatively, notice that ζ_p is a root of

$$x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$$

so since $\zeta_p \neq 1$, it follows that ζ_p is a root of

$$x^{p-1} + x^{p-2} + \cdots + x + 1$$

Plugging in ζ_p gives the result. □

We now extend the previous lemma to a "twisted" version where we use ζ_p^a in place of ζ_p.

Lemma 6.3.3. *Let $p \in \mathbb{N}^+$ be an odd prime and let $a \in \mathbb{Z}$.*

- *If $p \mid a$, then $\sum_{k=0}^{p-1} \zeta_p^{ak} = p$.*
- *If $p \nmid a$, then $\sum_{k=0}^{p-1} \zeta_p^{ak} = 0$.*

Proof. If $p \mid a$, then $\zeta_p^{ak} = (\zeta_p^a)^k = 1^k$ for all k, so

$$\sum_{k=0}^{p-1} \zeta_p^{ak} = \sum_{k=0}^{p-1} 1 = p$$

Suppose that $p \nmid a$. We then have $\zeta_p^a \neq 1$, so

$$\begin{aligned}
\sum_{k=0}^{p-1} \zeta_p^{ak} &= \sum_{k=0}^{p-1} (\zeta_p^a)^k \\
&= 1 + \zeta_p^a + (\zeta_p^a)^2 + \cdots + (\zeta_p^a)^{p-1} \\
&= \frac{(\zeta_p^a)^p - 1}{\zeta_p^a - 1} \\
&= \frac{1 - 1}{\zeta_p^a - 1} \\
&= 0
\end{aligned}$$

Alternatively, notice that the function $\psi\colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ defined by $\psi(x) = \overline{a} \cdot x$ is a bijection (because $\overline{a} \in U(\mathbb{Z}/p\mathbb{Z})$ has an inverse). Thus, every element in the list $0, a, 2a, 3a, \ldots, (p-1)a$ is congruent modulo p to exactly one element in the list $0, 1, 2, 3, \ldots, p-1$ and we can use the previous lemma. □

Definition 6.3.4. *Let $p \in \mathbb{N}^+$ be an odd prime and let $a \in \mathbb{Z}$. The* quadratic Gauss sum *(relative to a) is*

$$G_a = \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \cdot \zeta_p^{ak}$$

We let

$$G = G_1 = \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \cdot \zeta_p^{k}$$

Lemma 6.3.5. *Let $p \in \mathbb{N}^+$ be an odd prime. For any $a \in \mathbb{Z}$, we have $G_a = (\frac{a}{p}) \cdot G$.*

Proof. Suppose first that $p \mid a$. We then have $\zeta_p^{ak} = 1$ for all k, so

$$\begin{aligned}
G_a &= \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \cdot \zeta_p^{ak} \\
&= \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \\
&= 0
\end{aligned}$$

where the last line follows from the fact that exactly half of the elements in $\{1, 2, \ldots, p-1\}$ are quadratic residues. The result follows from the fact that $(\frac{a}{p}) = 0$ whenever $p \mid a$ by definition.

Suppose that $p \nmid a$. Notice that the function $\psi\colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ defined by $\psi(x) = \overline{a} \cdot x$ is a bijection (because $\overline{a} \in U(\mathbb{Z}/p\mathbb{Z})$ has an inverse). Thus, the every element in the list $0, a, 2a, 3a, \ldots, (p-1)a$ is congruent modulo p to exactly one element in the list $0, 1, 2, 3, \ldots, p-1$. It follows that

$$\begin{aligned}\left(\frac{a}{p}\right) \cdot G_a &= \left(\frac{a}{p}\right) \cdot \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \cdot \zeta_p^{ak} \\ &= \sum_{k=0}^{p-1} \left(\frac{ak}{p}\right) \cdot \zeta_p^{ak} \\ &= \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \cdot \zeta_p^{k} \\ &= G\end{aligned}$$

Now $(\frac{a}{p}) = \pm 1$, so multiplying both sides by $(\frac{a}{p})$ gives the result. □

Theorem 6.3.6. *Let $p \in \mathbb{N}^+$ be an odd prime. We have*

$$G^2 = \begin{cases} p & \text{if } p \equiv 1 \pmod 4 \\ -p & \text{if } p \equiv 3 \pmod 4 \end{cases}$$

Proof. We evaluate the sum $\sum_{a=0}^{p-1} G_a \cdot G_{-a}$ in two different ways. We have $G_0 \cdot G_0 = 0 \cdot 0 = 0$ and if $p \nmid a$ then

$$\begin{aligned}G_a \cdot G_{-a} &= \left(\frac{a}{p}\right) \cdot G \cdot \left(\frac{-a}{p}\right) \cdot G \\ &= \left(\frac{-1}{p}\right) \cdot \left(\frac{a^2}{p}\right) \cdot G^2 \\ &= \left(\frac{-1}{p}\right) \cdot G^2\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{a=0}^{p-1} G_a \cdot G_{-a} &= \sum_{a=1}^{p-1} G_a \cdot G_{-a} \\ &= \sum_{a=1}^{p-1} \left(\frac{-1}{p}\right) \cdot G^2 \\ &= \left(\frac{-1}{p}\right) \cdot (p-1) \cdot G^2\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\sum_{a=0}^{p-1} G_a \cdot G_{-a} &= \sum_{a=0}^{p-1} \left(\left(\sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \cdot \zeta_p^{ak} \right) \cdot \left(\sum_{\ell=0}^{p-1} \left(\frac{\ell}{p}\right) \cdot \zeta_p^{-a\ell} \right) \right) \\
&= \sum_{a=0}^{p-1} \sum_{k=0}^{p-1} \sum_{\ell=0}^{p-1} \left(\frac{k}{p}\right) \left(\frac{\ell}{p}\right) \cdot \zeta_p^{a(k-\ell)} \\
&= \sum_{k=0}^{p-1} \sum_{\ell=0}^{p-1} \left(\left(\frac{k}{p}\right) \left(\frac{\ell}{p}\right) \cdot \left(\sum_{a=0}^{p-1} \zeta_p^{a(k-\ell)} \right) \right) \\
&= \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \left(\frac{k}{p}\right) \cdot p \\
&= \sum_{k=1}^{p-1} p \\
&= p(p-1)
\end{aligned}$$

Therefore, we have

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

from which we conclude that

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

Using the fact that $(\frac{-1}{p}) = 1$ when $p \equiv 1 \pmod 4$ and $(\frac{-1}{p}) = -1$ when $p \equiv 3 \pmod 4$ completes the proof. □

Now that we have found a sum of roots of unity equal that squares to $\pm p$, we are ready for the proof of Quadratic Reciprocity.

Proof of Quadratic Reciprocity. Let p and q be distinct odd primes. Let G be the quadratic Gauss sum for p, i.e.

$$G = \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \cdot \zeta_p^k$$

Let

$$p^* = \begin{cases} p & \text{if } p \equiv 1 \pmod 4 \\ -p & \text{if } p \equiv 3 \pmod 4 \end{cases}$$

By Euler's Criterion, we know that

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

We now work in $\mathbb{Z}[\zeta_p]$. Using the fact that $G^2 = p^*$ by the previous theorem, we conclude that

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

and hence

$$G^q \equiv \left(\frac{p^*}{q}\right) \cdot G \pmod q$$

Now using the fact that q is an odd prime, we also have

$$\begin{aligned} G^q &\equiv \left(\sum_{k=0}^{p-1}\left(\frac{k}{p}\right)\cdot\zeta_p^k\right)^q \pmod{q} \\ &\equiv \sum_{k=0}^{p-1}\left(\frac{k}{p}\right)^q\cdot\zeta_p^{qk} \pmod{q} \\ &\equiv \sum_{k=0}^{p-1}\left(\frac{k}{p}\right)\cdot\zeta_p^{qk} \pmod{q} && \text{(since } q \text{ is odd)} \\ &\equiv G_q \pmod{q} \\ &\equiv \left(\frac{q}{p}\right)\cdot G \pmod{q} \end{aligned}$$

It follows that

$$\left(\frac{p^*}{q}\right)\cdot G \equiv \left(\frac{q}{p}\right)\cdot G \pmod{q}$$

and multiplying both sides by G we conclude that

$$\left(\frac{p^*}{q}\right)\cdot p^* \equiv \left(\frac{q}{p}\right)\cdot p^* \pmod{q}$$

This congruence also hold in $\mathbb{Z}$ because every element of $\mathbb{Z}[\zeta_p]$ is an algebraic integer. Using the fact that $\gcd(p^*, q) = 1$, we conclude that

$$\left(\frac{p^*}{q}\right) \equiv \left(\frac{q}{p}\right) \pmod{q}$$

Since $q \geq 3$, it follows that

$$\left(\frac{p^*}{q}\right) = \left(\frac{q}{p}\right)$$

If $p \equiv 1 \pmod 4$, then $p^* = p$ and we are done. Suppose that $p \equiv 3 \pmod 4$ so that $p^* = -p$. We then have

$$\left(\frac{p^*}{q}\right) = \left(\frac{-1}{q}\right)\cdot\left(\frac{p}{q}\right)$$

If $q \equiv 1 \pmod 4$, then $(\frac{-1}{q}) = 1$ and we are done. If $q \equiv 3 \pmod 4$, then $(\frac{-1}{q}) = -1$ and we are done. □

Example 6.3.7. *Compute*

$$\left(\frac{-42}{61}\right)$$

Solution. We have

$$\begin{aligned}
\left(\frac{-42}{61}\right) &= \left(\frac{-1}{61}\right)\cdot\left(\frac{2}{61}\right)\cdot\left(\frac{3}{61}\right)\cdot\left(\frac{7}{61}\right) \\
&= 1\cdot\left(\frac{2}{61}\right)\cdot\left(\frac{3}{61}\right)\cdot\left(\frac{7}{61}\right) && (\text{since } 61\equiv 1 \pmod 4) \\
&= 1\cdot(-1)\cdot\left(\frac{3}{61}\right)\cdot\left(\frac{7}{61}\right) && (\text{since } 61\equiv 5 \pmod 8) \\
&= 1\cdot(-1)\cdot\left(\frac{61}{3}\right)\cdot\left(\frac{61}{7}\right) && (\text{since } 61\equiv 1 \pmod 4) \\
&= 1\cdot(-1)\cdot\left(\frac{1}{3}\right)\cdot\left(\frac{5}{7}\right) \\
&= 1\cdot(-1)\cdot 1\cdot\left(\frac{5}{7}\right) \\
&= 1\cdot(-1)\cdot 1\cdot\left(\frac{7}{5}\right) && (\text{since } 5\equiv 1 \pmod 4) \\
&= 1\cdot(-1)\cdot 1\cdot\left(\frac{2}{5}\right) \\
&= 1\cdot(-1)\cdot 1\cdot(-1) && (\text{since } 5\equiv 5 \pmod 8) \\
&= 1
\end{aligned}$$

Therefore -42 is a quadratic residue modulo 61. □

Proposition 6.3.8. *Let p be prime. We have 3 is a quadratic residue modulo p if and only if one of the following is true.*

- $p = 2$
- $p \equiv 1 \pmod{12}$
- $p \equiv 11 \pmod{12}$

Proof. First notice that $3 \equiv 1 \pmod 2$, so 3 is clearly a quadratic residue modulo 2. Also we have that $3 \equiv 0 \pmod 3$, so 3 is not a quadratic residue modulo 3. Suppose then that $p > 3$. By Quadratic Reciprocity and the fact that $3 \equiv 3 \pmod 4$, we know that

$$\left(\frac{3}{p}\right) = \begin{cases} \left(\frac{p}{3}\right) & \text{if } p \equiv 1 \pmod 4 \\ -\left(\frac{p}{3}\right) & \text{if } p \equiv 3 \pmod 4 \end{cases}$$

We now have the following cases.

- Suppose that $p \equiv 1 \pmod 3$ and $p \equiv 1 \pmod 4$. We then have
$$\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right) = \left(\frac{1}{3}\right) = 1$$
- Suppose that $p \equiv 1 \pmod 3$ and $p \equiv 3 \pmod 4$. We then have
$$\left(\frac{3}{p}\right) = -\left(\frac{p}{3}\right) = -\left(\frac{1}{3}\right) = 1$$

- Suppose that $p \equiv 2 \pmod 3$ and $p \equiv 1 \pmod 4$. We then have

$$\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right) = \left(\frac{2}{3}\right) = -1$$

- Suppose that $p \equiv 2 \pmod 3$ and $p \equiv 3 \pmod 4$. We then have

$$\left(\frac{3}{p}\right) = -\left(\frac{p}{3}\right) = -\left(\frac{2}{3}\right) = -(-1) = 1$$

Thus, 3 is a quadratic residue modulo p exactly in the first and fourth cases. Using the Chinese Remainder Theorem, this is equivalent to either $p \equiv 1 \pmod{12}$ or $p \equiv 11 \pmod{12}$. □

Chapter 7

Dedekind Domains and Factorizations of Ideals

7.1 Ideals as Missing Elements

Let R be an integral domain, and consider the set $\mathcal{H}$ of all ideals of R. We obtain a function $f\colon R \to \mathcal{H}$ by defining $f(a) = \langle a \rangle$, i.e. every element of R is sent to the ideal that it generates. Notice that given $a, b \in R$, we have that $f(a) = f(b)$ if and only if a and b are associates in R. Hence f fails to be injective whenever $U(R) \supsetneq \{1\}$ (which is almost always since $-1 \in U(R)$ and $1 \neq -1$ if R has characteristic other than 2). Also, notice that f is surjective exactly when R is a PID.

Applying f to an element of R results in a principal ideal that consists of the set of all multiples of a. If we pass from elements to principal ideals in this way, we would like to still be able to multiply elements. Fortunately, we worked out the basics about multiplication of ideals in Homework 1.

Definition 7.1.1. *Let R be a commutative ring and let I and J be ideals of R. We define*

$$IJ = \{c_1d_1 + c_2d_2 + \cdots + c_kd_k : k \in \mathbb{N}^+, c_i \in I, d_i \in J\}$$

In other words, IJ is the additive subgroup of R generated by the set $\{cd : c \in I, d \in J\}$.

Proposition 7.1.2. *If I and J are ideals of a commutative ring R, then IJ is an ideal of R.*

Proof. Notice that $0 \in IJ$ because $0 \in I$, $0 \in J$, and $0 \cdot 0 = 0$. Also, the set IJ is clearly closed under addition because if we take two elements of IJ, say

$$a_1b_1 + a_2b_2 + \cdots + a_\ell b_\ell$$

and

$$c_1d_1 + c_2d_2 + \cdots + c_kd_k$$

where $a_i, c_i \in I$ and $b_id_i \in J$, then

$$a_1b_1 + a_2b_2 + \cdots + a_\ell b_\ell + c_1d_1 + c_2d_2 + \cdots + c_kd_k$$

is of the correct form so is an element of IJ. Suppose then that we have an element $c_1d_1+c_2d_2+\cdots+c_kd_k \in IJ$ where $c_i \in I$ and $d_i \in J$ and that $r \in R$. We have

$$\begin{aligned} r(c_1d_1 + c_2d_2 + \cdots + c_kd_k) &= r(c_1d_1) + r(c_2d_2) + \cdots + r(c_kd_k) \\ &= (rc_1)d_1 + (rc_2)d_2 + \ldots (rc_k)d_k \end{aligned}$$

Now since $r \in R$ and each $c_i \in I$, we know that each $rc_i \in I$ because I is an ideal of R. Therefore, the last expression above witnesses the fact that $r(c_1d_1 + c_2d_2 + \cdots + c_kd_k) \in IJ$. Thus, the set IJ is closed under multiplication by any element of R. □

We also had the following result on Homework 1.

Proposition 7.1.3. *Let R be a commutative ring and let I and J be ideals of R. Suppose that $I = \langle a \rangle$ and $J = \langle b \rangle$. We then have $IJ = \langle ab \rangle$.*

Proof. We first show that $IJ \subseteq \langle ab \rangle$. Consider an element $c_1d_1 + c_2d_2 + \cdots + c_kd_k \in IJ$ where $c_i \in I$ and $d_i \in J$. Since each $c_i \in I = \langle a \rangle$, we may fix $r_i \in R$ with $c_i = r_ia$. Since each $d_i \in J = \langle b \rangle$, we may fix $s_i \in R$ with $d_i = s_ib$. We then have

$$\begin{aligned} c_1d_1 + c_2d_2 + \cdots + c_kd_k &= (r_1a)(s_1b) + (r_2a)(s_2b) + \cdots + (r_ka)(s_kb) \\ &= (r_1s_1)(ab) + (r_2s_2)(ab) + \cdots + (r_ks_k)(ab) \\ &= (r_1s_1 + r_2s_2 + \cdots + r_ks_k)ab \end{aligned}$$

Since $r_1s_1 + r_2s_2 + \cdots + r_ks_k \in R$, it follows that $c_1d_1 + c_2d_2 + \cdots + c_kd_k \in \langle ab \rangle$. Therefore, $IJ \subseteq \langle ab \rangle$.

We now show that $\langle ab \rangle \subseteq IJ$. Let $x \in \langle ab \rangle$ and fix $r \in R$ with $x = r(ab)$. We then have $x = (ra)b$ and since $ra \in \langle a \rangle = I$ and $b \in \langle b \rangle = J$, it follows that $x \in IJ$. Therefore, $\langle ab \rangle \subseteq IJ$.

Combining the above two facts, we conclude that $IJ = \langle ab \rangle$. □

The previous proposition tells us the following. Suppose that $a, b \in R$. If we multiply in R and then form $f(ab)$, we obtain the same thing as first computing $f(a)$ and $f(b)$, and then multiplying the corresponding ideals. In other words, f preserves multiplication (if would be a homomorphism except for the fact that R under multiplication of elements, and $\mathcal{H}$ under multiplication of ideals, are not groups). We also know that given $p \in R$, we have that p is a prime element of R if and only if $\langle p \rangle$ is a prime ideal of R, so our map also preserve "primeness". In other words, this transition from elements to ideals loses a bit of information by identifying associates, but respects the important algebraic aspects of multiplication.

In fact, it turns out that this identification of associates, although a loss in information, is actually helpful in stating results more elegantly. For example, consider $R = \mathbb{Z}$. When thinking about factorizations of elements, we considered the factorizations $6 = 2 \cdot 3$ and $6 = (-2) \cdot (-3)$ as the "same" because we could pair off elements up to associates. Thus, in this case, the presence of distinct associates made a fundamental result less elegant. However, consider the fact that as ideals, these factorizations become

$$\langle 6 \rangle = \langle 2 \rangle \cdot \langle 3 \rangle \qquad \langle 6 \rangle = \langle -2 \rangle \cdot \langle -3 \rangle$$

What has been gained? The answer is that although $2 \neq -2$ and $3 \neq -3$, we do have that $\langle 2 \rangle = \langle -2 \rangle$ and $\langle 3 \rangle = \langle -3 \rangle$. Thus, as ideals these factorization are in fact exactly the same. We do still have to cope with order because $\langle 6 \rangle = \langle 3 \rangle \cdot \langle 2 \rangle$, but this switch from elements to ideals results in a more elegant statement.

The above ideas are particularly elegant in a ring like $R = \mathbb{Z}[i]$ where there are more units besides ± 1. Recall that 2 is not irreducible in R and that

$$2 = (1 + i)(1 - i) = (-i) \cdot (1 + i)^2$$

Notice that initially it is not obvious that the terms in first factorization are associates. To make the terms equal and realize that a square is present, we needed to introduce the unit out front in the latter factorization. Compare this to the situation using ideals where we simply have

$$\langle 2 \rangle = \langle 1 + i \rangle \cdot \langle 1 - i \rangle \qquad \langle 2 \rangle = \langle 1 + i \rangle \cdot \langle 1 + i \rangle$$

without having to explicitly introduce the unit (notice that $\langle -i \rangle = R$ because $-i$ is a unit of R, and $RI = I$ for every ideal so we need not include it). Also we do in fact have $\langle 1 + i \rangle \cdot = \langle 1 - i \rangle$ as ideals, so each of these factorization are actually identical.

So far, we have be working in PIDs R where the function $f\colon R \to \mathcal{H}$ is surjective. How should we interpret the situation when R is not a PID and so there are elements of $\mathcal{H}$ that are not in range(f). Now by passing from elements on R side to ideals on the $\mathcal{H}$ side we have introduced new objects. What do they represent?

The key insight is the following. Let $a \in R$ and consider the ideal $I = \langle a \rangle$. By definition, we have that I is the set of multiples of a. In other words, principal ideals are exactly the set of multiples of some element. If we abstract away and consider a set J of elements of R, what properties should it have if it is to be considered the set of multiples of some element? Certainly we should have $0 \in J$ because 0 is a multiple of everything. If you take two multiples of an element and add them, you should end up with another multiple of the element, so J should be closed under addition (and similarly subtraction). Finally, if you take a multiple of an element and multiply it by an arbitrary $r \in R$, then you should up with another multiple of the element, so J should be closed under multiplication by any element of R. Looking at the properties we just listed, we see that they are precisely the defining properties of an ideal!

The creative leap now is to view any ideal I, principal or not, as the set of multiples of some "element". The scare quotes are there because if I is a nonprincipal ideal, then I is not literally the set of multiples of any element of R. But if we take that leap we can try to imagine that there is a magical "ideal" element that is not really an element of R whose multiples in R are precisely the elements of I. Thus, the ideal is the "shadow" of this nonexistent element inside of the ring R. When viewed this way, ideals serve the role of completing R be including elements should be there may not actually exist.

This is all rather cute and a bit mind-bending, but the real question is whether taking this point of view actually allows us to do something insightful and allow us to build an interesting theory. Consider the number field $K = \mathbb{Q}(\sqrt{-5})$ and its ring of integers $R = \mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$ (since $-5 \not\equiv 1 \pmod 4$). We know from earlier that R is not a UFD because we have the two factorizations

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

where each of the four factors are irreducible in R and none are associates (since $U(R) = \{\pm 1\}$). We also showed that although 2 is an irreducible element of R, it is not a prime element because $2 \mid (1+\sqrt{-5})(1-\sqrt{-5})$ but $2 \nmid 1 + \sqrt{-5}$ and $2 \nmid 1 - \sqrt{-5}$. When we analyzed this situation in the past, we threw our hands in the air. What more could we do? We can't factor 2 any further so can not rectify the situation.

Since R is not a UFD, it is not a PID, and hence there are ideals that are not the set of multiples of an element of R. In fact, on Homework 8, we show that $P = \langle 2, 1 + \sqrt{-5} \rangle$ is a nonprincipal prime ideal of R. Thus, P does not equal the set of multiples of any element of $\mathcal{O}_K$. However, try to think about P as the set of multiples of some magical unicorn "ideal" element that just happens not to live in R. Since $2 \in I$ and $1 + \sqrt{-5}$ are in P, each of these elements are multiples of this "ideal" prime element. Thus, we have found a new "element" other than ± 1 that is a common divisor of 2 and $1 + \sqrt{-5}$. Perhaps we can use this element to refine the above factorization.

Before getting into the calculations, we need a simple result about how to multiply nonprincipal but finitely generated ideals. It extends the above the result about multiplication of principal ideals.

Proposition 7.1.4. *Let R be a commutative ring and let I and J be ideals of R. Suppose that $I = \langle a_1, a_2, \ldots, a_k \rangle$ and $J = \langle b_1, b_2, \ldots, b_\ell \rangle$. We then have $IJ = \langle \{a_i b_j : 1 \leq i \leq k, 1 \leq j \leq \ell\} \rangle$. In other words, IJ is generated as an ideal by all products $a_i b_j$.*

Proof. The proof is essentially the same as the one for principal ideals, and the only difficulty is working with the notation. □

Let's see this in practice, Working in $R = \mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$, we had the two different factorizations of 6:

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

As above, consider the nonprincipal prime ideal $P = \langle 2, 1+\sqrt{-5}\rangle$. Let's examine what happens when we multiply this ideal by itself to form $P^2 = PP$. By the above proposition, we know that

$$\begin{aligned} P^2 &= \langle 2, 1+\sqrt{-5}\rangle \cdot \langle 2, 1+\sqrt{-5}\rangle \\ &= \langle 4, 2+2\sqrt{-5}, 2+2\sqrt{-5}, -4+2\sqrt{-5}\rangle \\ &= \langle 4, 2+2\sqrt{-5}, -4+2\sqrt{-5}\rangle \end{aligned}$$

Now 2 divides each of these three generators in R, so 2 divides every elements of P^2 in R. It follows that $P^2 \subseteq \langle 2\rangle$. Now

$$6 = (2+2\sqrt{-5}) - (-4+2\sqrt{-5}) \in P^2$$

and hence $2 = 6 - 4 \in P^2$. From this we conclude that $\langle 2\rangle \subseteq P^2$. Combining these two containments, it follows that

$$P^2 = \langle 2\rangle$$

Thus, although 2 is not prime but also does not factor nontrivially in R, the ideal $\langle 2\rangle$ factors as the square of an "ideal" element. One could also consider the ideal $\langle 2, 1-\sqrt{-5}\rangle$ to think about a new common divisor of these elements. However, a simple calculation shows that $\langle 2, 1-\sqrt{-5}\rangle = \langle 2, 1+\sqrt{-5}\rangle = P$ (use the fact that $2 \cdot \sqrt{-5}$ is in both ideals).

We can do the same thing with other pairs. Let

$$Q = \langle 3, 1+\sqrt{-5}\rangle \qquad\qquad L = \langle 3, 1-\sqrt{-5}\rangle$$

As in the case for P, one can check that Q are L are nonprincipal maximal ideals of R (they both have index 3 in R), and hence both are nonprincipal prime ideals of R. In contrast to the case with 2 in place of 3, one can also check that $Q \neq L$. Now

$$\begin{aligned} QL &= \langle 3, 1+\sqrt{-5}\rangle \cdot \langle 3, 1-\sqrt{-5}\rangle \\ &= \langle 9, 3+3\sqrt{-5}, 3-3\sqrt{-5}, 6\rangle \\ &= \langle 3\rangle \end{aligned}$$

so we have succeeded in factoring 3 into prime ideal elements. One can also check that

$$\begin{aligned} PQ &= \langle 2, 1+\sqrt{-5}\rangle \cdot \langle 3, 1+\sqrt{-5}\rangle \\ &= \langle 6, 2+2\sqrt{-5}, 3+3\sqrt{-5}, -4+2\sqrt{-5}\rangle \\ &= \langle 1+\sqrt{-5}\rangle \end{aligned}$$

and

$$\begin{aligned} PL &= \langle 2, 1-\sqrt{-5}\rangle \cdot \langle 3, 1-\sqrt{-5}\rangle \\ &= \langle 6, 2-2\sqrt{-5}, 3-3\sqrt{-5}, -4-2\sqrt{-5}\rangle \\ &= \langle 1-\sqrt{-5}\rangle \end{aligned}$$

Let's put this all together. Starting with the two bad factorizations

$$2 \cdot 3 = 6 = (1+\sqrt{-5})(1-\sqrt{-5})$$

we obtain think in terms of ideals to see the factorizations

$$\langle 2\rangle \cdot \langle 3\rangle = \langle 6\rangle = \langle 1+\sqrt{-5}\rangle \cdot \langle 1-\sqrt{-5}\rangle.$$

Now we insert our new ideal factorizations to turn these into

$$P^2 \cdot QL = \langle 6 \rangle = PQ \cdot PL$$

Now ideal multiplication is commutative associative (see below), so these factorization as exactly the same thing after rearranging! By passing from elements to ideals we have recovered uniqueness of factorization into primes (where here we mean prime ideals rather than prime elements), at least in this one small case. This is extremely exciting!

In hindsight, let's compare the factorizations

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

in $\mathbb{Z}[\sqrt{-5}]$ with the factorizations

$$4 \cdot 15 = 60 = 6 \cdot 10$$

in $\mathbb{Z}$. The latter does not violate unique factorization because it was not actually a factorization into primes. We can break down $4 = 2^2$, $15 = 3 \cdot 5$, $6 = 2 \cdot 3$, and $10 = 2 \cdot 5$ to see that

$$(2 \cdot 2) \cdot (3 \cdot 5) = 60 = (2 \cdot 3) \cdot (2 \cdot 5)$$

At first sight, the factorization stymied us because we could not break down the elements any further. However, by passing to ideals we succeeded in writing

$$PP \cdot QL = \langle 6 \rangle = PQ \cdot PL$$

just as in $\mathbb{Z}$.

The rest of this chapter is an elaboration of these ideas that will ultimately demonstrate that this is the "correct" setting in which to work. However, we must travel a long road to get to our destination. We want to replace elements by ideals, but immediately this raises several questions. Does multiplication of ideals behave in a similar way to multiplication of elements? The following result is straightforward.

Proposition 7.1.5. *Let R be an integral domain.*

1. *$R \cdot I = I$ for all ideals I of R.*
2. *$I \cdot J = J \cdot I$ for all ideals I and J of R.*
3. *$I \cdot (J \cdot K) = (I \cdot J) \cdot K$ for all ideals I, J, K or R.*

Proof. The first is immediate from the observation that $1 \in R$.

For the second, notice that $I \cdot J$ is the additive subgroup of R generated by the set $\{cd : c \in I, d \in J\}$ while $J \cdot I$ is the additive subgroup of R generated by the set $\{da : d \in J, c \in I\}$. These two sets are equal because R is commutative, so $I \cdot J = J \cdot I$.

For the third, check that each side is the additive subgroup of R generated by the set $\{abc : a \in I, b \in J, c \in K\}$. The notation is horrible, but the argument is straightforward. □

Therefore, multiplication of ideals is commutative and associative (just like multiplication of elements), and $R = \langle 1 \rangle$ serves as a multiplicative identity. However, it takes much more work to extend other fundamental properties of element multiplication to ideal multiplication. For example, we would really hope that the following are true:

- If $I \cdot J = I \cdot K$ and $I \neq \{0\}$, then $J = K$
- If $I \subseteq J$, then there exists K such that $I = JK$

The latter arises from the idea that if $I \subseteq J$, then all multiples of the ideal element I are multiples of the ideal element J, so that might suggest that J "divides" I. In other words, it is the abstract analogue of the fact that if $\langle a \rangle \subseteq \langle b \rangle$, then $b \mid a$. We will eventually prove these properties in certain nice rings, but they are far from obvious. The following alternate characterization of prime ideals in terms of ideals is not too difficult and will be extremely useful.

Proposition 7.1.6. *Let R be integral domain. Let P be a proper ideal of R. The following are equivalent.*

1. *P is a prime ideal of R, i.e. whenever $a, b \in R$ are such that $ab \in P$, either $a \in P$ or $b \in P$.*
2. *Whenever I and J are ideals of R such that $IJ \subseteq P$, either $I \subseteq P$ or $J \subseteq P$.*

Proof. $1 \to 2$: Suppose that P is a prime ideal of R. Let I and J be ideal of R and assume that $IJ \subseteq P$. Suppose that $I \not\subseteq P$, and fix $a \in I \backslash P$. We show that $J \subseteq P$. Let $b \in J$. We then have that $ab \in IJ$, so since we are assuming that $IJ \subseteq P$, we know that $ab \in P$. Now P is prime ideal, so either $a \in P$ or $b \in P$. Since a was chosen to be an element of $I \backslash P$, we must have $b \in P$. Since $b \in J$ was chosen arbitrarily, we conclude that $J \subseteq P$.

$2 \to 1$: Suppose we know 2. Let $a, b \in R$ and assume that $ab \in P$. Define $I = \langle a \rangle$ and $J = \langle b \rangle$. We know from above that $IJ = \langle ab \rangle$. Since $ab \in P$ and P is an ideal, it follows that $IJ \subseteq P$. Thus, by 2, either $I \subseteq P$ or $J \subseteq P$. If $I \subseteq P$, then $a \in P$, while if $J \subseteq P$, then $b \in P$. □

The grand hope is that in "nice" rings, every ideal factors uniquely (up to order) as a product of prime ideals. This is our ultimate goal for the rings we have been studying.

7.2 Dedekind Domains

As we have seen, it is not generally true that $\mathcal{O}_K$ is always a PID. However, for any number field K, the ring $\mathcal{O}_K$ always has several pleasing ring theoretic properties. The wider class of rings that belong to are called Dedekind domains.

Definition 7.2.1. *A* Dedekind domain *is an integral domain R with the following three properties:*

- *R is Noetherian.*
- *Every nonzero prime ideal of R is a maximal ideal.*
- *R is integrally closed in its field of fractions. In other words, if we let F be the field of fractions of R, then whenever $\alpha \in F$ is a root of a monic polynomial in $R[x]$, we must have $\alpha \in R$.*

The above definition requires some time and experience to understand and appreciate. Before proving that some of more exotic rings we have studied are Dedekind domains, we first prove that nice rings we have studied have the above properties.

Theorem 7.2.2. *Every PID is a Dedekind domain.*

Proof. Let R be a PID. Corollary 3.3.6 tells us that every R is Noetherian (remember that Noetherian is equivalent to the statement that every ideal is finitely generated). Corollary 3.2.13 tells us that every nonzero prime ideal in R is a maximal ideal (recall that these ideals are precisely the ones generated by irreducible/prime elements).

For the third property, we generalize the proofs of the Rational Root Theorem and Corollary 4.5.10 to R. Let F be the field of fractions of R. Let $\alpha \in F$ be a root of a monic polynomial $p(x) \in R[x]$. Write

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

where each $a_i \in R$. Write $\alpha = \frac{b}{c}$ where $b, c \in R$ with $c \neq 0$ and where b and c are relatively prime in R (this is possible because R is a PID and hence greatest common divisor exist). We then have

$$(b/c)^n + a_{n-1} \cdot (b/c)^{n-1} + \cdots + a_1 \cdot (b/c) + a_0 = 0$$

Multiplying through by c^n we get

$$b^n + a_{n-1}b^{n-1}c + \cdots + a_1bc^{n-1} + a_0c^n = 0$$

From this, we see that

$$b^n = c \cdot [-(a_{n-1}b^{n-1} + \cdots + a_1bc^{n-2} + a_0c^{n-1})]$$

and hence $c \mid b^n$ in R. From this we conclude that any irreducible/prime divisor of c in R would be a divisor of b^n and hence a divisor of b. Since b and c are relatively prime in R, they have no common irreducible/prime divisors in R, and hence c must not be divisible by any irreducible/prime in R. However, we know that every nonzero nonunit element of R is divisible by an irreducible element, so the only possible conclusion is that c is unit in R. Thus, $\alpha = b/c = bc^{-1} \in R$. It follows that R is integrally closed in F. □

Let K be a number field. We prove on the homework that the field of fractions of $\mathcal{O}_K$ is K. Thus, we must ask whether $\mathcal{O}_K$ is integrally closed in K. Now by definition, $\mathcal{O}_K$ is the set of algebraic integers of K, which is the set of elements of K that are roots of monic polynomials in $\mathbb{Z}[x]$. Since we have thrown in all elements of K that roots of monic polynomials, it is natural to believe that $\mathcal{O}_K$ is integrally closed in K. However, the definition only dealt with polynomials having coefficients from $\mathbb{Z}$. We now have to consider monic polynomials having coefficients in $\mathcal{O}_K$ that are not in $\mathbb{Z}[x]$. Are roots of these polynomials necessarily in $\mathcal{O}_K$ as well? The question is similar to the following analytic idea. Suppose that you take the closure of a set A to obtain a set B. Now what happens when you take the closure of B? It is natural to hope that the new closure is B again, but this is not obvious because there are more points in B and hence the possibility of new limit points.

For example, consider $K = \mathbb{Q}(\zeta_3)$. By definition of $\mathcal{O}_K$, every element of K that is a root of some monic polynomial with integer coefficients lies in $\mathcal{O}_K$. However, consider the polynomial

$$p(x) = x^3 + \zeta_3 x^2 + (5 - 2\zeta_3)x + 7$$

If there is a root α of $p(x)$ that lies in K, it is not at all obvious that this root is an algebraic integer (and hence belongs to $\mathcal{O}_K$) because it is far from clear how to find a monic polynomial in $\mathbb{Z}[x]$ that has α as a root.

Theorem 7.2.3. *Let K be a number field. We then have have that $\mathcal{O}_K$ is integrally closed in its field of fractions K.*

Proof. We know from the homework that K is the field of fractions of $\mathcal{O}_K$. Let $p(x) \in \mathcal{O}_K[x]$ be a monic polynomial. Write

$$p(x) = x^n + \beta_{n-1}x^{n-1} + \cdots + \beta_1 x + \beta_0$$

where each $\beta_i \in \mathcal{O}_K$. Let $\alpha \in K$ be a root of $p(x)$.

Since each $\beta_i \in \mathcal{O}_K$, we know that each β_i is an algebraic integer. Using Theorem 4.5.6 that $\mathbb{Z}[\beta_i]$ is finitely generated as an additive abelian group. In fact, letting m_i be the degree of some monic polynomial $f_i(x)$ $in \mathbb{Z}[x]$ having β_i as a root, the ring $\mathbb{Z}[\beta_i]$ is finitely generated as an additive abelian group by the set $\{\beta_i^k : 0 \leq k < m_i\}$ (intuitively we can reduce powers of β_i above m_i to lower powers with integer coefficients using the monic polynomial $f_i(x)$). Using this, the ring $\mathbb{Z}[\beta_0, \beta_1, \ldots, \beta_{n-1}]$ is finitely generated as an additive abelian group by the finite set

$$\{\prod_{i=0}^{n-1} \beta_i^{k_i} : 0 \leq i \leq n-1, 0 \leq k_i \leq m_i\}$$

because again we can reduce larger powers of β_i above m_i to lower powers using $f_i(x)$.

One can now show that the ring $\mathbb{Z}[\beta_0, \beta_1, \ldots, \beta_{n-1}, \alpha]$ is finitely generated as an additive abelian group because we can use that fact that

$$\alpha^n = -(\beta_{n-1}\alpha^{n-1} + \cdots + \beta_1\alpha + \beta_0)$$

to reduce large powers of α beyond n to smaller powers of α with coefficients from the β_i. Thus, the ring $\mathbb{Z}[\beta_0, \beta_1, \ldots, \beta_{n-1}, \alpha]$ is generated as an additive abelian group by the finite set

$$\{\alpha^\ell \cdot \prod_{i=0}^{n-1} \beta_i^{k_i} : 0 \leq i \leq n-1, 0 \leq k_i \leq m_i, 0 \leq \ell < n\}$$

Since α is an element of some subring of $\mathbb{C}$ that is finitely generated as an additive abelian group, we may use Theorem 4.5.6 to conclude that α is an algebraic integer. Since $\alpha \in K$, it follows by definition that $\alpha \in \mathcal{O}_K$. Therefore, $\mathcal{O}_K$ is integrally closed in K. □

Lemma 7.2.4. *Let K be a quadratic number field. For any nonzero ideal I of $\mathcal{O}_K$, we have that $\mathbb{N}^+ \cap I \neq \emptyset$.*

Proof. Let I be a nonzero ideal of $\mathcal{O}_K$. Fix $\alpha \in I$ with $\alpha \neq 0$. Notice that $N(\alpha) = \overline{\alpha} \cdot \alpha \in I$. We have that $N(\alpha) \in \mathbb{Z}$ and also that $N(\alpha) \neq 0$ because $\alpha \neq 0$ (and hence $\overline{\alpha} \neq 0$ as well). If $N(\alpha) > 0$, we are done. If not $N(\alpha) < 0$, notice that $-N(\alpha) \in\in \mathbb{N}^+ \cap I$ as well. □

Theorem 7.2.5. *Let K be a quadratic number field. For any nonzero ideal I of $\mathcal{O}_K$, the ring $\mathcal{O}_K/I$ is finite.*

Proof. Fix a square-free $d \in \mathbb{Z}\backslash\{1\}$ with $K = \mathbb{Q}(\sqrt{d})$. Let I be a nonzero ideal of $\mathcal{O}_K$. By the previous lemma, we may fix $m \in \mathbb{N}^+ \cap I$ and let $J = \langle m \rangle$. We then have that $J \subseteq I$, so to show that $\mathcal{O}_K/I$ is finite it suffices to show that $\mathcal{O}_K/J$ is finite (because if $\alpha + J = \beta + J$, then $\alpha - \beta \in J \subseteq I$, so $\alpha + I = \beta + I$). Let

$$\eta = \begin{cases} \sqrt{d} & \text{if } d \not\equiv 1 \pmod 4 \\ \frac{1+\sqrt{d}}{2} & \text{if } d \equiv 1 \pmod 4 \end{cases}$$

Thus, in either case, we have

$$\mathcal{O}_K = \mathbb{Z}[\eta] = \{a + b\eta : a, b \in \mathbb{Z}\}$$

We claim that every element of $\mathcal{O}_K/J$ is represented by an element of the form $r_1 + r_2\eta$ with $0 \leq r_i < m$. To see this, let $a, b \in \mathbb{Z}$ be arbitrary. Since $m \in \mathbb{N}^+$, we may fix $q_i, r_i \in \mathbb{Z}$ with $a = q_1 m + r_1$, $b = q_2 m + r_2$ and $0 \leq r_i < m$. We then have that

$$\begin{aligned} (a + b\eta) - (r_1 + r_2\eta) &= (a - r_1) + (b - r_2)\eta \\ &= q_1 m + q_2 m\eta \\ &= m \cdot (q_1 + q_2\eta) \end{aligned}$$

Thus, $(a + b\eta) - (r_1 + r_2\eta) \in J$ and hence $(a + b\eta) + J = (r_2 + r_2\eta) + J$. It follows that every element of $\mathcal{O}_K/J$ is represented by an element of the form $r_1 + r_2\eta$ with $0 \leq r_i < m$. Since there are finitely many such elements, we conclude that $\mathcal{O}_K/J$ is finite. As mentioned above, this implies that $\mathcal{O}_K/I$ is finite. □

Corollary 7.2.6. *Let K be a quadratic number field. We then have that $\mathcal{O}_K$ is Noetherian and that every nonzero prime ideal of $\mathcal{O}_K$ is maximal. Thus, $\mathcal{O}_K$ is a Dedekind domain.*

Proof. We first show that $\mathcal{O}_K$ is Noetherian. Below, we use the notation $(G : H)$ to mean the index of the subgroup H inside the group G (that is, the number of cosets of H in G). Suppose that

$$I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \dots$$

is a strictly increasing sequence of ideals of $\mathcal{O}_K$. Notice that $I_2 \neq \{0\}$, so by the previous theorem we know that $\mathcal{O}_K/I_2$ is finite, i.e. that $(\mathcal{O}_K : I_2)$ is finite (where we are viewing each of these as additive groups). Since $I_2 \subsetneq I_3$, we know that $(I_3 : I_2) \geq 2$, so since

$$(\mathcal{O}_K : I_2) = (\mathcal{O}_K : I_3) \cdot (I_3 : I_2)$$

it follows that $(\mathcal{O}_K : I_3) < (\mathcal{O}_K : I_2)$. Following this argument, we see that $(\mathcal{O}_K : I_{m+1}) < (\mathcal{O}_K : I_m)$ for all m. However, this is a contradiction because there does not exist an infinite decreasing sequence of natural numbers.

Suppose now that P is a nonzero prime ideal of $\mathcal{O}_K$. We then know that $\mathcal{O}_K/P$ is an integral domain. By the previous corollary, we also know that $\mathcal{O}_K/P$ is finite. Since every finite integral domain is a field, we conclude that $\mathcal{O}_K/P$ is a field, and hence that P is a maximal ideal.

Combining both of these with the fact that $\mathcal{O}_K$ is integrally closed in its field of fractions K (from above), we conclude that $\mathcal{O}_K$ is a Dedekind domain. □

In fact, one can prove that $\mathcal{O}_K$ is a Dedekind domain for every number field K (not just the quadratic ones). To do this, it suffices to prove Lemma 7.2.4 and Theorem 7.2.5 for general number fields K because that is all we used in the previous corollary. In order to prove Lemma 7.2.4, one can define a norm function on $\mathcal{O}_K$ that has similar properties, but this requires some more advanced field theory and Galois-theoretic ideas. For Theorem 7.2.5, it suffices to show that $\mathcal{O}_K$ is always a finitely generated abelian group. In fact, one can show that $\mathcal{O}_K$ always has an *integral basis*, i.e. a finite set $\alpha_1, \alpha_2, \dots, \alpha_n$ that is linearly independent over $\mathbb{Q}$ and such that

$$\mathcal{O}_K = \{k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n : k_i \in \mathbb{Z}\}$$

We were able to do this in the quadratic case by using our knowledge of the elements of $\mathcal{O}_K$ to realize that $\{1, \eta\}$ was an integral basis. In general, this is significantly harder to prove in a general $\mathcal{O}_K$ because it is not always easy to get one's hands on the elements. However, it is still true, but requires some more tools that we currently have at our disposal.

7.3 Factorizations of Ideals in Dedekind Domains

Lemma 7.3.1. *Let R be a Noetherian ring. Let $\mathcal{H}$ be a nonempty set of ideals of R. There exists a maximal element of $\mathcal{H}$, i.e. there exists $I \in \mathcal{H}$ such that there is no $J \in \mathcal{H}$ with $I \subsetneq J$.*

Proof. Suppose that $\mathcal{H}$ does not have a maximal element. Let $I_1 \in \mathcal{H}$ be arbitrary. Since I_1 is not a maximal element of $\mathcal{H}$, we may fix $I_2 \in \mathcal{H}$ with $I_1 \subsetneq I_2$. Continue to define a sequence of ideals recursively, i.e. given $I_n \in \mathcal{H}$, we know that I_n is not a maximal element of $\mathcal{H}$, so we may fix $I_{n+1} \in \mathcal{H}$ with $I_n \subsetneq I_{n+1}$. Thus, we have a constructed a chain of ideals

$$I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \dots$$

contradicting the fact that R is Noetherian. □

In the context of Dedekind domains, the next corollary is the analogue of the statement that every nonunit is divisible by some prime (where we are used ideal containment to represent divisibility). In the integers, this is proved by a simple induction. Notice how the Noetherian condition replaces the role of induction by forcing chains of ideals to end.

Corollary 7.3.2. *Let R be a Noetherian ring. Every proper ideal of R is contained in a maximal ideal of R.*

Proof. Let I be a proper ideal of R. Let $\mathcal{H}$ be the set of all proper ideals J of R such that $I \subseteq J$. Notice that $\mathcal{H} \neq \emptyset$ because $I \in \mathcal{H}$. A maximal element of $\mathcal{H}$ exists by the previous lemma, and is a maximal ideal of R containing I. □

Now that we know that every nontrivial ideal is "divisible" by some prime ideal, our next hope is to write every such ideal element as a product of prime ideals. As a small stepping stone to this, we first show that every nonzero ideal contains a product of nonzero prime ideals. Intuitively, an ideal with few members corresponds to a large "ideal element" (since the ideal is thought of as a set of multiples, and larger elements have fewer multiples). Thus, this next proposition is saying that no ideal is so "large" when viewed as an element (or equivalently has so few members) that it can not be subsumed by a product of primes. Again, the key idea is to use the Noetherian property to take the place of an inductive proof.

Proposition 7.3.3. *Let R be a Dedekind domain and let I be a nonzero ideal of R. There exist nonzero prime ideals $P_1, P_2, \ldots, P_k$ of R (not necessarily distinct) with $P_1 P_2 \cdots P_k \subseteq I$.*

Proof. Let $\mathcal{H}$ be the set of all nonzero ideals I that do not contain a product of nonzero prime ideals. Suppose that $\mathcal{H} \neq \emptyset$. Let I be a maximal element of $\mathcal{H}$. Notice that I is certainly not prime itself, so we may fix $a, b \in R$ with $ab \in I$ but $a \notin I$ and $b \notin I$. Consider the ideals $I + \langle a \rangle$ and $I + \langle b \rangle$. Each of these ideals properly contain I, so by maximality, each of these ideals contains a product of nonzero primes. Fix nonzero prime ideals P_i and Q_j such that $P_1 P_2 \cdots P_k \subseteq I + \langle a \rangle$ and $Q_1 Q_2 \cdots Q_\ell \subseteq I + \langle b \rangle$.

We claim that $P_1 P_2 \cdots P_k Q_1 Q_2 \cdots Q_\ell \subseteq I$. To see this, it suffices to show that $cd \in I$ whenever $c \in P_1 P_2 \cdots P_k$ and $d \in Q_1 Q_2 \cdots Q_\ell$ (because the product is the set of sums of these elements). Let $c \in P_1 P_2 \cdots P_k$ and let $d \in Q_1 Q_2 \cdots Q_\ell$. We then have $c \in I + \langle a \rangle$, so we may fix $x \in I$ and $r \in R$ with $c = x + ra$. We also have $d \in I + \langle b \rangle$, so we may fix $y \in I$ and $s \in R$ with $d = y + sb$. We then have

$$cd = (x + ra)(y + sb) = xy + ray + sbx + rsab$$

Now $x, y \in I$ and $ab \in I$, so $cd \in I$. Thus, $P_1 P_2 \cdots P_k Q_1 Q_2 \cdots Q_\ell \subseteq I$. We have found a product of nonzero ideals that is a subset of I, so we have arrived at a contradiction. Therefore, $\mathcal{H} = \emptyset$, and the result follows. □

Proposition 7.3.4. *Let R be a Dedekind domain and let I be a proper nonzero ideal of R. Let F be the field of fractions of R. There exists $\gamma \in F \backslash R$ such that $\gamma I \subseteq R$, i.e. such that $\gamma a \in R$ for all $a \in I$.*

Proof. Since $I \neq \{0\}$, we may fix a nonzero $d \in I$. By the previous proposition, we know that $\langle d \rangle$ contains a product of nonzero prime ideals, and so we may fix nonzero prime ideals $P_1, P_2, \ldots, P_k$ such that $P_1 P_2 \cdots P_k \subseteq \langle d \rangle$ where k is as small as possible. Since I is a proper ideal, we may fix a maximal ideal M containing I. Maximal ideals are always prime, so M is a prime ideal. We have

$$P_1 P_2 \cdots P_k \subseteq \langle d \rangle \subseteq I \subseteq M$$

so by Proposition 7.1.6, we must have $P_i \subseteq M$ for some i. Relabeling if necessary, we may assume that $P_1 \subseteq M$. Now R is a Dedekind domain, so since P_1 is a nonzero prime ideal, it must be maximal. Hence, $M = P_1$ and we have

$$P_1 P_2 \cdots P_k \subseteq \langle d \rangle \subseteq I \subseteq P_1$$

Since k was chosen as small as possible, we know that $P_2 \cdots P_k \not\subseteq \langle d \rangle$, hence we may fix $c \in P_2 \cdots P_k$ such that $c \notin \langle d \rangle$. We then have that $d \nmid c$ in R, so $\frac{c}{d} \in F \backslash R$. Now for any $a \in I$, we have $a \in P_1$, so $ac \in P_1 P_2 \cdots P_k \subseteq \langle d \rangle$. and hence $d \mid ac$ in R. Thus, $\frac{c}{d} \cdot a = \frac{ac}{d} \in R$ for all $a \in I$, and we may take $\gamma = \frac{c}{d} \in F \backslash R$. □

Lemma 7.3.5. *Let R be an integral domain. Let I be an ideal of R and let $d \in R$. Assume that $d \mid a$ for all $a \in I$. Let*

$$J = \frac{1}{d} \cdot I = \{r \in R : rd \in I\}$$

We then have that J is an ideal of R and $I \subseteq J$.

Proof. We clearly have $0 \in J$ since $0 \cdot d = 0 \in I$. If $a, b \in J$, then $ad \in I$ and $bd \in I$, hence $(a+b)d = ad + bd \in I$, and hence $a + b \in J$. If $a \in J$ and $r \in R$, then $ad \in I$, so $(ra)d = r(ad) \in I$ and hence $ra \in J$. Thus, J is an ideal of R.

Finally if $a \in I$, then $ad = da \in I$, and hence $a \in J$. Thus, $I \subseteq J$. □

Lemma 7.3.6. *Let R be a commutative ring and let I be an ideal of R. For any $d \in R$, we have $\langle d \rangle \cdot I = \{da : a \in I\}$.*

Proof. We clearly have $\{da : a \in I\} \subseteq \langle d \rangle \cdot I$. We show that reverse containment. Let $x \in \langle d \rangle \cdot I$. Fix $b_1, b_2, \ldots, b_n \in \langle d \rangle$ and $a_1, a_2, \ldots, a_n \in I$ such that $x = b_1a_1 + b_2a_2 + \cdots + b_na_n$. For each i, fix $r_i \in R$ with $b_i = r_id$. We then have

$$\begin{aligned} x &= b_1a_1 + b_2a_2 + \cdots + b_na_n \\ &= (r_1d)a_1 + (r_2d)a_2 + \cdots + (r_nd)a_n \\ &= d \cdot (r_1a_1 + r_2a_2 + \ldots r_na_n) \end{aligned}$$

Since $r_1a_1 + r_2a_2 + \ldots r_na_n \in I$, it follows that $x \in \{da : a \in I\}$. Thus, $\langle d \rangle \cdot I \subseteq \{da : a \in I\}$. □

Theorem 7.3.7. *Let R be a Dedekind domain. For all ideals I of R, there exists a nonzero ideal J of R such that IJ is principal. In fact, if $I \neq \{0\}$ and d is an arbitrary nonzero element of I, then there exists a nonzero ideal J of R such that $IJ = \langle d \rangle$.*

Proof. If $I = \{0\}$, this is trivial by taking $J = R = \langle 1 \rangle$, so assume that $I \neq \{0\}$. Fix $d \in I$ with $d \neq 0$. Define

$$\begin{aligned} J &= \{r \in R : \langle r \rangle \cdot I \subseteq \langle d \rangle\} \\ &= \{r \in R : d \mid ra \text{ for all } a \in I\} \end{aligned}$$

Notice that J is an ideal of R (check this) and $J \neq \{0\}$ because $d \in J$. We certainly have $IJ \subseteq \langle d \rangle$, and we claim that $IJ = \langle d \rangle$. Since $IJ \subseteq \langle d \rangle$, we know that $d \mid x$ for all $x \in IJ$, so using Lemma 7.3.5 we see that the set

$$K = \frac{1}{d} \cdot IJ = \{r \in R : rd \in IJ\}$$

is an ideal of R. If $K = R$, then $1 \in K$, so $d \in IJ$ and hence $IJ = \langle d \rangle$ which completes the proof.

Suppose then that K is a proper ideal of R. Let F be the field of fractions of R. By the above proposition, we may fix $\gamma \in F \backslash R$ with $\gamma K \subseteq R$. Notice that if $c \in J$, then $cd \in IJ$ because $d \in I$, hence $c \in K$. Thus, $J \subseteq K$. It follows that $\gamma J \subseteq \gamma K \subseteq R$.

We now claim that $\gamma J \subseteq J$. Let $c \in J$. Since $\gamma J \subseteq R$, notice that $\gamma c \in R$. We need to show that $\gamma c \in J$, so we need to show that $d \mid (\gamma c) \cdot a$ for all $a \in I$. Let $a \in I$. We have $ca = ac \in IJ$, so $d \mid ac$. It follows that $\frac{ac}{d} \in K$, and hence $\gamma \cdot \frac{ac}{d} \in R$. Thus, $d \mid (\gamma c) \cdot a$.

Since J is an ideal of R and R is Noetherian, we may fix $b_1, b_2, \ldots, b_n$ such that $J = \langle b_1, b_2, \ldots, b_n \rangle$. Now $\gamma J \subseteq J$, so for each i, we have $\gamma b_i \in J$ and hence may fix $r_{i,j} \in R$ with

$$\gamma b_i = r_{i,1}b_1 + r_{i,2}b_2 + \cdots + r_{i,n}b_n$$

Let M be the $n \times n$ matrix $M = [r_{i,j}]$ viewed as a matrix over the field F, and let $\mathbf{v}$ be the $n \times 1$ column vector $\mathbf{v} = [b_i]$. The above equation simply says that $\alpha \mathbf{v} = M\mathbf{v}$, and notice that $\mathbf{v} \neq \mathbf{0}$ because $J \neq \{0\}$.

Thus, γ is an eigenvalue of M. Let $f(x) = \det(xI - M)$ be the characteristic polynomial of M. Notice that $f(x)$ is a monic polynomial (of degree n) and that $f(x) \in R[x]$ because all entries in M are elements of R. Since γ is an eigenvalue of A, we know that γ is a root of the characteristic polynomial $f(x)$. Now R is Dedekind domain, so R is integrally closed in F and hence we must have $\gamma \in R$. This is a contradiction, and hence it is not possible that K is a proper ideal of R. □

Corollary 7.3.8. *Let R be a Dedekind domain. Let I, J, K be ideals of R. If $IJ = IK$ and $I \neq \{0\}$, then $J = K$.*

Proof. Suppose that $IJ = IK$ and $I \neq \{0\}$. Fix a nonzero ideal L such that LI is principal. Fix $d \in R$ with $LI = \langle d \rangle$ and notice that $d \neq 0$ because both I and L are nonzero ideals. Multiplying both sides of $IJ = IK$ by L and using the fact that ideal multiplication is associative, we see that $\langle d \rangle \cdot J = \langle d \rangle \cdot K$.

We now claim that $J = K$. Suppose first that $a \in J$. We then have $da \in \langle d \rangle \cdot J$, so $da \in \langle d \rangle \cdot K$. Thus, we may fix $b \in K$ with $da = db$. Since R is a an integral domain and $d \neq 0$, we may cancel to conclude that $a = b \in K$. Thus, $J \subseteq K$. Similarly, we have $K \subseteq J$. Combining these, we conclude that $J = K$. □

Corollary 7.3.9. *Let R be a Dedekind domain. Let I and J be ideals of R. We have If $I \subseteq J$ if and only if there exists an ideal K such that $I = JK$*

Proof. Suppose first that there exists an ideal K with $I = JK$. Since $JK \subseteq J$, we clearly have $I \subseteq J$.

Suppose conversely that $I \subseteq J$. First assume that J is principal and fix $d \in R$ with $J = \langle d \rangle$. We then have $I \subseteq \langle d \rangle$, so $d \mid a$ for all $a \in I$. By Lemma 7.3.5, the set

$$K = \frac{1}{d} \cdot I = \{r \in R : rd \in I\}$$

is an ideal of R. We then have that $JK = \langle d \rangle \cdot K = I$ (check this last equality).

Suppose then that $I \subseteq J$ but J is nonprincipal. Fix a nonzero ideal L of R such that JL is principal. We then have $IL \subseteq JL$, so as JL is principal, we may fix an ideal K with $IL = JLK$. Since $L \neq \{0\}$, we may use the previous corollary to conclude that $I = JK$. □

Theorem 7.3.10. *Let R be a Dedekind domain.*

- *Every nonzero proper ideal of R can be written as a product of prime ideals.*
- *If $P_1, P_2, \ldots, P_k, Q_1, Q_2, \ldots, Q_\ell$ are nonzero primes ideals of R with $P_1 P_2 \cdots P_k = Q_1 Q_2 \cdots Q_\ell$, then $k = \ell$ and there exists $\sigma \in S_k$ such that $P_i = Q_{\sigma(i)}$ for all i.*

Therefore, every nonzero proper ideal of R factors uniquely as a product of prime ideals.

Proof. Let $\mathcal{H}$ be the collection of all nonzero proper ideals of R that can *not* be written as a product of prime ideals. Suppose that $\mathcal{H} \neq \emptyset$. Since R is Noetherian, we may fix a maximal element I of $\mathcal{H}$. Now I is a proper ideal of R, so since R is Noetherian we may fix a maximal ideal M of R with $I \subseteq M$. By the previous corollary, we may fix an ideal K of R with $I = MK$. We then have that $I \subseteq K$. Also, if $I = K$, then $RI = I = MI$, so $R = M$ by cancellation, which is a contradiction. Thus, $I \subsetneq K$, so $K \notin \mathcal{H}$ because I is a maximal element of $\mathcal{H}$. Therefore, K is a product of nonzero prime ideals, and since M is also a prime ideal (because it is a maximal ideal), it follows that $I = MK$ can be written as a product of prime ideals. This is a contradiction, so we must have $\mathcal{H} = \emptyset$.

Suppose that $P_1, P_2, \ldots, P_k$ and $Q_1, Q_2, \ldots, Q_\ell$ are nonzero primes ideals of R with $P_1 P_2 \cdots P_k = Q_1 Q_2 \cdots Q_\ell$. Since $P_1 P_2 \cdots P_k \subseteq P_1$, we then have

$$Q_1 Q_2 \cdots Q_\ell \subseteq P_1$$

Since P_1 is prime, Lemma 7.1.6 implies that there exists j with $Q_j \subseteq P_1$. Now R is a Dedekind domain, so Q_j is a maximal ideal and hence $P_1 = Q_j$. We may now cancel this factor from both sides and repeat. At

no stage can we have a nonempty product of prime ideals equal to R, so eventually we pair up all the P_i with a Q_j and leave nothing left over. □

Theorem 7.3.11. *If R is a Dedekind domain, then R is a PID if and only if R is a UFD.*

Proof. We know that PIDs are always UFDs in general. Suppose then that R is a Dedekind domain that is also a UFD. Since R is a UFD, we know that every nonzero nonunit element of R can be written as a product of irreducibles, and we also know that every irreducible element of R is prime. Combining these, we conclude that every nonzero nonunit element of R can be written as a product of prime elements.

Suppose now that I is an arbitrary ideal of R. We need to show that I is principal. If either $I = \{0\}$ or $I = R$, then R is trivially principal, so assume that I is a nonzero proper ideal of R. We know that we may fix a nonzero ideal L of R such IL is principal. If $L = R$, then $IL = I$, and hence I is principal. Suppose then that L is a nonzero proper ideal of R. Notice that IL is a nonzero proper ideal of R, so we may fix a nonzero nonunit $a \in R$ with $IL = \langle a \rangle$. From above, we may write $a = p_1 p_2 \cdots p_n$ where the p_i are prime elements of R. We then have

$$IL = \langle a \rangle = \langle p_1 \rangle \cdot \langle p_2 \rangle \cdots \langle p_n \rangle$$

Since R is a Dedekind domain, we may fix nonzero prime ideals Q_i and R_j such that $I = Q_1 Q_2 \cdots Q_k$ and $L = R_1 R_2 \cdots R_\ell$. We then have

$$Q_1 Q_2 \cdots Q_k R_1 R_2 \cdots R_\ell = \langle p_1 \rangle \cdot \langle p_2 \rangle \cdots \langle p_n \rangle$$

Now each ideal $\langle p_i \rangle$ is a prime ideal because it is generated by a prime element, so by unique factorization of ideals into primes, we know that each Q_i equals some $\langle p_j \rangle$. Thus, I is a product of principal prime ideals, and hence is principal. □

7.4 The Class Group

Definition 7.4.1. *Let R be an integral domain. Define a relation $\sim$ on the set of nonzero ideals of R by saying that $I \sim J$ if there exist nonzero $a, b \in R$ with $\langle a \rangle \cdot I = \langle b \rangle \cdot J$.*

Thus, two ideals are equivalent if they can be made equal via multiplication by (nonzero) principal ideals.

Proposition 7.4.2. *Let R be an integral domain. The above relation $\sim$ is an equivalence relation on the set of nonzero ideals of R.*

Proof. We clearly have $\langle 1 \rangle \cdot I = \langle 1 \rangle \cdot I$ for any nonzero ideal I of R, so $\sim$ is symmetric. Also, $\sim$ is symmetric immediately from the definition. Suppose that I and J are nonzero ideals of R with $I \sim J$ and $J \sim K$. Fix nonzero $a, b, c, d \in R$ with $\langle a \rangle \cdot I = \langle b \rangle \cdot J$ and $\langle c \rangle \cdot J = \langle d \rangle \cdot K$. We then have

$$\begin{aligned} \langle ca \rangle \cdot I &= \langle c \rangle \cdot \langle a \rangle \cdot I \\ &= \langle c \rangle \cdot \langle b \rangle \cdot J \\ &= \langle b \rangle \cdot \langle c \rangle \cdot J \\ &= \langle b \rangle \cdot \langle d \rangle \cdot K \\ &= \langle bd \rangle \cdot J \end{aligned}$$

Notice that $ca \neq 0$ and $bd \neq 0$ because R is an integral domain, so $I \sim K$. Therefore, $\sim$ is an equivalence relation. □

The following lemma says that if R is a subring of $\mathbb{C}$ (such as when $R = \mathcal{O}_K$ for a number field K), then $I \sim J$ exactly when I and J have the same "shape" when viewed as a subset of $\mathbb{C}$.

Lemma 7.4.3. *Let R be a subring of $\mathbb{C}$. Let I and J be nonzero ideals of R. We then have that $I \sim J$ if and only if there exists a nonzero $\lambda \in \mathbb{C}$ with $\lambda \cdot I = J$ (where $\lambda \cdot I = \{\lambda c : c \in I\}$).*

Proof. Suppose first that $I \sim J$. Fix nonzero $a, b \in R$ with $\langle a \rangle \cdot I = \langle b \rangle \cdot J$. We claim that $\frac{a}{b} \cdot I = J$.

- We first show that $\frac{a}{b} \cdot I \subseteq J$. Let $c \in I$. We have $ac \in \langle a \rangle \cdot I$, so we must have $ac \in \langle b \rangle \cdot J$. By Lemma 7.3.6, we may fix $d \in J$ with $ac = bd$. We then have
$$\frac{a}{b} \cdot c = \frac{ac}{b} = \frac{bd}{b} = d$$
so $\frac{a}{b} \cdot c \in J$.
- We now show that $J \subseteq \frac{a}{b} \cdot I$. Let $d \in J$. We have $bd \in \langle b \rangle \cdot J$, so we must have $bd \in \langle a \rangle \cdot I$. By Lemma 7.3.6, we may fix $c \in I$ with $bd = ac$. We then have
$$d = \frac{bd}{b} = \frac{ac}{b} = \frac{a}{b} \cdot c$$
so $d \in \frac{a}{b} \cdot I$.

Combining the two containments, we conclude that $\frac{a}{b} \cdot I = J$. Thus, we may let $\lambda = \frac{a}{b} \in \mathbb{C}$.

Suppose conversely that there exists a nonzero $\lambda \in \mathbb{C}$ with $\lambda \cdot I = J$. Since I is a nonzero ideal of R, we may fix $b \in I$ with $b \neq 0$. We then have $\lambda b \in J$, so we may fix $a \in J$ with $\lambda b = a$. Notice that $a \neq 0$ because $\lambda \neq 0$ and $b \neq 0$. We claim that $\langle a \rangle \cdot I = \langle b \rangle \cdot J$.

- We first show $\langle a \rangle \cdot I \subseteq \langle b \rangle \cdot J$. Let $x \in \langle a \rangle \cdot I$. By Lemma 7.3.6, we may fix $c \in I$ with $x = ac$. Since $\lambda \cdot I = J$, we have $\lambda c \in J$. Now $x = ac = \lambda bc = b \cdot (\lambda c)$, so $x \in \langle b \rangle \cdot J$.
- We now show that $\langle b \rangle \cdot J \subseteq \langle a \rangle \cdot I$. Let $y \in \langle b \rangle \cdot J$. By Lemma 7.3.6, we may fix $d \in J$ with $y = bd$. Since $J = \lambda \cdot I$, we may fix $c \in I$ with $d = \lambda c$. Now $y = bd = b(\lambda c) = (\lambda b) \cdot c = ac$, so $y \in \langle a \rangle \cdot I$.

Combining the two containments, we conclude that $\langle a \rangle \cdot I = \langle b \rangle \cdot J$. □

Proposition 7.4.4. *Let R be an integral domain. For any nonzero ideal I of R, we have $R \sim I$ if and only if I is a principal ideal. Thus $\overline{R}$ is the set of all principal ideals of R.*

Proof. Suppose first that I is a nonzero principal ideal of R. Fix $a \in I$ such that $I = \langle a \rangle$. We then have that $\langle a \rangle \cdot R = \langle a \rangle = I = \langle 1 \rangle \cdot I$, so $R \sim I$.

Suppose conversely that I is a nonzero ideal of R and that $R \sim I$. Fix nonzero elements $a, b \in R$ with $\langle a \rangle \cdot R = \langle b \rangle \cdot I$. We then have $\langle a \rangle = \langle b \rangle \cdot I$. Thus $a \in \langle b \rangle \cdot I$, so using Lemma 7.3.6 we may fix $c \in I$ with $a = bc$. We claim that $I = \langle c \rangle$. Since $c \in I$, we clearly have $\langle c \rangle \subseteq I$. Let $x \in I$. We then have $bx \in \langle b \rangle \cdot I$, so $bx \in \langle a \rangle$. Fix $r \in R$ with $bx = ra$. We then have $bx = rbc$, so since R is an integral domain and $b \neq 0$, we must have $x = rc$ Thus, $x \in \langle c \rangle$. It follows that $I = \langle c \rangle$ and hence I is a principal ideal of R. □

By the previous proposition, every nonzero principal ideal of R has the same "shape" as R. In the case where $R = \mathbb{Z}[i]$ is the Gaussian integers, we have that R is a PID and hence all nonzero ideals of R look just like R except for scaling and rotation (as we saw in Section 3.6). However, the situation is much more interesting when R is not a PID. Consider the case where $R = \mathcal{O}_{\mathbb{Q}(\sqrt{-5})} = \mathbb{Z}[\sqrt{-5}]$. We still have that all principal ideals of R have the sam shape as R, but we know that nonprincipal ideals of R exist. Let $I = \langle 2, 1 + \sqrt{-5} \rangle$ and recall from the homework that I is nonprincipal. Drawing the elements of R as a subset of $\mathbb{C}$ gives a geometric demonstration of these differences. Consider the case where $J = \langle 3, 1 + \sqrt{-5} \rangle$. One can show that J is also a nonprincipal ideal (by the similar argument). Notice that

$$\begin{aligned} \langle 1 + \sqrt{-5} \rangle \cdot I &= \langle 1 + \sqrt{-5} \rangle \cdot \langle 2, 1 + \sqrt{-5} \rangle \\ &= \langle 2 + 2\sqrt{-5}, -4 + 2\sqrt{5} \rangle \end{aligned}$$

and

$$\begin{aligned}\langle 2\rangle \cdot J &= \langle 2\rangle \cdot \langle 3, 1+\sqrt{-5}\rangle \\ &= \langle 6, 2+2\sqrt{-5}\rangle\end{aligned}$$

Now

$$-4+2\sqrt{5} = (2+2\sqrt{-5}) - 6 \in \langle 6, 2+2\sqrt{-5}\rangle$$

and

$$6 = (2+2\sqrt{-5}) - (-4+2\sqrt{5}) \in \langle 2+2\sqrt{-5}, -4+2\sqrt{5}\rangle$$

so

$$\langle 1+\sqrt{-5}\rangle \cdot I = \langle 2\rangle \cdot J$$

and hence $I \sim J$. In fact, one can show that any two nonprincipal ideals of $R = \mathbb{Z}[\sqrt{-5}]$ are equivalent and hence $\sim$ has exactly two equivalence classes in this case.

Proposition 7.4.5. *Let R be an integral domain and let I_1, I_2, J_1, J_2 be nonzero ideals of R. If $I_1 \sim I_2$ and $J_1 \sim J_2$, then $I_1 J_1 \sim I_2 J_2$.*

Proof. Since $I_1 \sim I_2$, we may fix nonzero $a_1, a_2 \in R$ with $\langle a_1\rangle \cdot I_1 = \langle a_2\rangle \cdot I_2$. Since $J_1 \sim J_2$, we may fix nonzero $b_1, b_2 \in R$ with $\langle b_1\rangle \cdot J_1 = \langle b_2\rangle \cdot J_2$. We then have

$$\begin{aligned}\langle a_1 b_1\rangle \cdot I_1 J_1 &= \langle a_1\rangle \cdot \langle b_1\rangle \cdot I_1 \cdot J_1 \\ &= \langle a_1\rangle \cdot I_1 \cdot \langle b_1\rangle \cdot J_1 \\ &= \langle a_2\rangle \cdot I_2 \cdot \langle b_2\rangle \cdot J_2 \\ &= \langle a_2\rangle \cdot \langle b_2\rangle \cdot I_2 \cdot J_2 \\ &= \langle a_2 b_2\rangle \cdot I_2 J_2\end{aligned}$$

Since $a_1 b_1 \neq 0$ and $a_2 b_2 \neq 0$ (because R is an integral domain), it follows that $I_1 J_1 \sim I_2 J_2$. □

Theorem 7.4.6. *Let R be an Dedekind domain. Define a binary operation on the equivalence classes of $\sim$ by letting $\overline{I} \cdot \overline{J} = \overline{IJ}$. Under this operation, the set of equivalence classes of ideals of R becomes an abelian group.*

Proof. The previous proposition says that $\cdot$ is well-defined on the equivalence classes of $\sim$. Now associativity and commutativity of $\cdot$ follow immediately from the associativity and commutativity of ideal multiplication. Notice that for any nonzero ideal I of R, we have $RI = I = IR$ (since $1 \in R$), hence $\overline{R} \cdot \overline{I} = \overline{I} = \overline{I} \cdot \overline{R}$. Thus, $\overline{R}$ serves as an identity. Suppose that I is an arbitrary nonzero ideal of R. Since R is a Dedekind domain, Theorem 7.3.7 implies that we may fix a nonzero ideal J of R such that IJ is principal. We then have $R \sim IJ$ from above, so $\overline{I} \cdot \overline{J} = \overline{IJ} = \overline{R}$ (and hence $\overline{J} \cdot \overline{I} = \overline{R}$ by commutativity). Therefore, every element has an inverse. □

Definition 7.4.7. *Let R be a Dedekind domain. The above group is called the* (ideal) class group *of R.*

Notice that a Dedekind domain R is a UFD if and only if R is a PID if and only if the class group of R is trivial. Thus, one way to view the class group is that it provides a measure of how badly unique factorization and nonprincipality of ideals can fail. As mentioned above, the class group of $\mathbb{Z}[\sqrt{-5}]$ is a cyclic group of order 2. The following is a nontrivial theorem that is a fundamental result in algebraic number theory (and is beyond our scope).

Theorem 7.4.8. *If K be a number field, then the class group of $\mathcal{O}_K$ is always a finite group.*

Chapter 8

Cyclotomic Extensions

8.1 Cyclotomic Polynomials

Definition 8.1.1. *Let F be a field and let $n \in \mathbb{N}^+$. An* n^{th} root of unity *in F is an element $u \in F$ such that $u^n = 1$. In other words, an n^{th} root of unity is a root of the polynomial $x^n - 1 \in F[x]$. Working in a field F, we let R_n be the set of n^{th} roots of unity.*

Given any field F and any $n \in \mathbb{N}^+$, notice that F has at most n distinct n^{th} roots of unity because the polynomial $x^n - 1 \in F[x]$ has at most n distinct roots. However, in some fields there may be fewer. For example, 1 is the only 3^{rd} root of unity in $\mathbb{R}$, and ± 1 are the only fourth roots of unity in $\mathbb{R}$.

Working in $\mathbb{C}$, we know that $\zeta_n = e^{2\pi i/n}$ is an n^{th} root of unity. Furthermore, viewed as an element of $U(\mathbb{C})$, we have $|\zeta_n| = n$ (since $\zeta_n^n = 1$ but $\zeta_n^k = e^{2\pi ki/n} \neq 1$ whenever $0 < k < n$). Thus, ζ_n generates a subgroup of $U(\mathbb{C})$ with order n. Every element of this subgroup is an n^{th} root of unity, so $\mathbb{C}$ has at least n distinct n^{th} roots of unity (namely $\{1, \zeta_n, \zeta_n^2, \ldots, \zeta_n^{n-1}\}$). Therefore, $\mathbb{C}$ has exactly n distinct n^{th} roots of unity from above.

Proposition 8.1.2. *Let F be a field and let $n \in \mathbb{N}^+$. The set R_n is a cyclic subgroup of $U(F)$ with order equal to a divisor of n.*

Proof. Clearly $1 \in R_n$. If $u, w \in R_n$, then $u^n = 1 = w^n$, so $(uw)^n = u^n w^n = 1 \cdot 1 = 1$. If $u \in R_n$, then $u^n = 1$, so $u \neq 0$ and multiplying both sides by $(u^{-1})^n$ gives $1 = (u^{-1})^n$, so $u^{-1} \in R_n$.

We know that $|R_n| \leq n$ by the above comments, so R_n is a finite subgroup of $U(F)$. Therefore, R_n is cyclic by Theorem 2.6.1. Fix a generator w of R_n. We then have $w^n = 1$, so the order of w must be a divisor of n. Hence, $|R_n| = |w|$ is a divisor of n. □

Definition 8.1.3. *Let F be a field and let $n \in \mathbb{N}^+$. A* primitive n^{th} root of unity *in F is an element $u \in F$ such that $|u| = n$ when viewed as an element of $U(F)$. Working in a field F, we let P_n be the set of primitive n^{th} roots of unity.*

Using the previous proposition, notice that a primitive n^{th} root of unity exists in F exactly when $|R_n| = n$. We have that $\mathbb{C}$ has a primitive n^{th} root of unity for all n because $|\zeta_n| = n$ for all n. However, $\mathbb{R}$ does not have a primitive 4^{th} root of unity.

Suppose now that we are working in $\mathbb{C}$ and we have a fixed $n \in \mathbb{N}^+$. We know that primitive n^{th} roots of unity exist because ζ_n is one. Now the number of primitive n^{th} roots of unity is the number of generators of R_n. Since R_n is a cyclic group of order n, we know from Corollary 2.4.9 that R_n has exactly $\varphi(n)$ many generators, so $|P_n| = \varphi(n)$. Moreover, we know from Proposition 2.4.8 that ζ_n^k is a primitive n^{th} root of unity exactly when $\gcd(k, n) = 1$.

Consider now an arbitrary n^{th} root of unity u in a field F. We know that $u^n = 1$, so when viewed as an element of $U(F)$ we know that $|u|$ is a divisor of n. Suppose that $|u| = d$ for some $d \mid n$. Notice then that u is primitive d^{th} root of unity. Thus, every n^{th} root of unity is a primitive d^{th} root of unity for some $d \mid n$.

We want to understand the field extension $\mathbb{Q} \subseteq \mathbb{Q}(\zeta_n)$ or in general $\mathbb{Q} \subseteq \mathbb{Q}(\zeta_n^k)$. In order to do this, we first need to understand that minimal polynomial of ζ_n^k. Thus, we need to look at how $x^n - 1$ factors over $\mathbb{Q}$. We have the following.

- $x^2 - 1 = (x-1)(x+1)$
- $x^3 - 1 = (x-1)(x^2+x+1)$
- $x^4 - 1 = (x-1)(x+1)(x^2+1)$
- $x^5 - 1 = (x-1)(x^4+x^3+x^2+x+1)$
- $x^6 - 1 = (x-1)(x+1)(x^2+x+1)(x^2-x+1)$
- $x^7 - 1 = (x-1)(x^6+x^5+x^4+x^3+x^2+1)$
- $x^8 - 1 = (x-1)(x+1)(x^2+1)(x^4+1)$
- $x^9 - 1 = (x-1)(x^2+x+1)(x^6+x^3+1)$

It is straightforward to check that each of the above quadratics have no roots in $\mathbb{Q}$ and hence are irreducible in $\mathbb{Q}[x]$. It is harder to see that the other factors are irreducible in $\mathbb{Q}[x]$, but we will do this below. Looking at the the polynomials that arise, notice that

$$x^2 + x + 1 = (x - \zeta_3)(x - \zeta_3^2)$$

and

$$x^2 - x + 1 = (x - \zeta_6)(x - \zeta_6^5)$$

Thus, we are seeing the primitive n^{th} roots of unity appear as the roots of the various polynomials. We now turn this on its head by defining polynomials that have the primitive n^{th} roots of unity as the roots.

Definition 8.1.4. *Let $n \in \mathbb{N}^+$. Working in $\mathbb{C}$, let $S = \{k \in \{1, 2, \ldots, n\} : \gcd(k, n) = 1\}$ and define*

$$\begin{aligned}\Phi_n(x) &= \prod_{u \in P_n} (x - u) \\ &= \prod_{k \in S} (x - \zeta_n^k)\end{aligned}$$

As it stands, we only have $\Phi_n(x) \in \mathbb{C}[x]$ or better $\Phi_n(x) \in \mathbb{Q}(\zeta_n)[x]$. Also, since $|P_n| = \varphi(n)$, we have $\deg(\Phi_n(x)) = \varphi(n)$.

Suppose that $n \in \mathbb{N}^+$. The roots of $x^n - 1$ are all of the n^{th} roots of unity. Now any n^{th} root of unity is a primitive d^{th} root of unity for some unique $d \mid n$ as described above. Therefore

$$\begin{aligned}x^n - 1 &= \prod_{u \in R_n} (x - u) \\ &= \prod_{d \mid n} \prod_{u \in P_d} (x - u) \\ &= \prod_{d \mid n} \Phi_d(x)\end{aligned}$$

As we saw above, it turns out that several of the polynomials $\Phi_n(x)$ for small values of n are in $\mathbb{Z}[x]$, which is a bit surprising based only the definition. But we will see that this is no accident.

The above formula gives us a recursive method to calculate $\Phi_n(x)$. For example, to calculate $\Phi_5(x)$, we need only note that

$$x^5 - 1 = \Phi_1(x)\Phi_5(x) = (x-1)\Phi_5(x)$$

to conclude that

$$\Phi_5(x) = \frac{x^5-1}{x-1} = x^4 + x^3 + x^2 + x + 1$$

We can then use this to calculate $\Phi_{10}(x)$. We know that

$$x^{10} - 1 = \Phi_1(x)\Phi_2(x)\Phi_5(x)\Phi_{10}(x)$$

so

$$x^{10} - 1 = (x-1)(x+1)(x^4 + x^3 + x^2 + x + 1)\Phi_{10}(x)$$

It follows that

$$\Phi_{10}(x) = \frac{x^{10}-1}{(x^5-1)(x+1)} = \frac{x^5+1}{x+1} = x^4 - x^3 + x^2 - x + 1$$

Notice that if p is prime, then

$$x^p - 1 = \Phi_1(x)\Phi_p(x) = (x-1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$$

so

$$\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$$

We can now use this to compute $\Phi_{p^2}(x)$. We have

$$x^{p^2} - 1 = \Phi_1(x)\Phi_p(x)\Phi_{p^2}(x) = (x-1)(x^{p-1} + x^{p-2} + \cdots + 1)\Phi_{p^2}(x)$$

Since the product of the first two is $x^p - 1$, it follows that

$$\Phi_{p^2}(x) = \frac{x^{p^2}-1}{x^p-1} = x^{p(p-1)} + x^{p(p-2)} + \cdots + x^p + 1$$

as we saw in the case of $p^2 = 9$ above.

Lemma 8.1.5. *Let $F \subseteq K$ be a field extension. Let $f(x), g(x) \in F[x]$ and suppose that $g(x) \neq 0$. If $h(x) \in K[x]$ satisfies $f(x) = g(x)h(x)$, then in fact $h(x) \in F[x]$.*

Proof. Since $F[x]$ is a Euclidean domain, we may fix $q(x), r(x) \in F[x]$ with $f(x) = g(x)q(x) + r(x)$ and either $r(x) = 0$ or $\deg(r(x)) < \deg(g(x))$. This equation works in $K[x]$ too. Since quotients and remainders are unique in $K[x]$, and we also have $f(x) = g(x)h(x) + 0$, we must have $h(x) = q(x) \in K[x]$. □

Proposition 8.1.6. *$\Phi_n(x) \in \mathbb{Q}[x]$ and is monic for all $n \in \mathbb{N}^+$.*

Proof. The proof is by induction on n. It is trivial for $n = 1$. Suppose that the result is true for all $k < n$. We then have

$$x^n - 1 = \prod_{d|n} \Phi_d(x) = \Big(\prod_{d|n, d\neq n} \Phi_d(x)\Big) \cdot \Phi_n(x)$$

By induction the factor on the left (which is nonzero) is in $\mathbb{Q}[x]$ and we clearly have $x^n - 1 \in \mathbb{Q}[x]$. Thus, by the lemma, we have $\Phi_n(x) \in \mathbb{Q}[x]$. Also, induction tells us that the factor on the left is monic, so comparing leading terms it follows that $\Phi_n(x)$ is monic. □

Lemma 8.1.7. *Suppose that $f(x) \in \mathbb{Z}[x]$ and that $g(x), h(x) \in \mathbb{Q}[x]$ with $f(x) = g(x)h(x)$. Suppose further that both $f(x)$ and $g(x)$ are monic. We then have that $g(x), h(x) \in \mathbb{Z}[x]$ and that $h(x)$ is monic.*

Proof. By looking at leading terms, we conclude that $h(x)$ is also monic. By Gauss' Lemma, there exist $s, t \in \mathbb{Q}$ such that $s \cdot g(x) \in \mathbb{Z}[x]$, $t \cdot h(x) \in \mathbb{Z}[x]$, and

$$f(x) = (s \cdot g(x)) \cdot (t \cdot h(x))$$

Since $g(x)$ and $h(x)$ are monic, we must have $s, t \in \mathbb{Z}$. Now $f(x)$ is also monic, so looking at the leading term on the right we conclude that $st = 1$, so $s, t \in \{\pm 1\}$. It follows that $g(x), h(x) \in \mathbb{Z}[x]$. □

Theorem 8.1.8. *$\Phi_n(x) \in \mathbb{Z}[x]$ for all $n \in \mathbb{N}^+$.*

Proof. The proof is by induction on n. It is trivial for $n = 1$. Suppose that the result is true for all $k < n$. As above, we have

$$x^n - 1 = \prod_{d|n} \Phi_d(x) = (\prod_{d|n, d \neq n} \Phi_d(x)) \cdot \Phi_n(x)$$

By induction the factor on the left is a monic polynomial in $\mathbb{Z}[x]$ and we clearly have that $x^n - 1 \in \mathbb{Z}[x]$ is monic. Therefore, using the lemma, we conclude that $\Phi_n(x) \in \mathbb{Z}[x]$. □

Definition 8.1.9. *Let F be a field. We define a function $D\colon F[x] \to F[x]$, called the formal derivative, as follows. Given $f(x) \in F[x]$, say*

$$f(x) = \sum_{k=0}^{n} a_k x^k$$

we define

$$D(f(x)) = \sum_{k=1}^{n} k a_k x^{k-1} = \sum_{k=0}^{n-1} (k+1) a_{k+1} x^k$$

where we interpret $k \in F$ as $k = 1 + 1 + \cdots + 1$ (where there a total of k many 1's in the sum).

It is painful but straightforward to check that D obeys the usual derivative rules (such as the sum and product rules) for any field $F[x]$.

Proposition 8.1.10. *Let F be a field and let $f(x) \in F[x]$ be a nonzero polynomial. If $f(x)$ and $Df(x)$ are relatively prime in $F[x]$, then $ord_{q(x)}(f(x)) \leq 1$ for all irreducible $q(x) \in F[x]$.*

Proof. We prove the contrapositive. Suppose that $q(x) \in F[x]$ is irreducible and $ord_{q(x)}(f(x)) \geq 2$. Fix $g(x) \in F[x]$ with $f(x) = q(x)^2 g(x)$. We then have

$$\begin{aligned} Df(x) &= 2 \cdot q(x) \cdot Dq(x) \cdot g(x) + q(x)^2 \cdot Dg(x) \\ &= q(x) \cdot [2 \cdot Dq(x) \cdot g(x) + q(x) \cdot Dg(x)] \end{aligned}$$

so $q(x)$ is a nonunit common divisor of $f(x)$ and $Df(x)$. □

Corollary 8.1.11. *Let $n \in \mathbb{N}^+$ and let $p \in \mathbb{N}^+$ be prime with $p \nmid n$. Working in $\mathbb{Z}/p\mathbb{Z}[x]$, we have $ord_{q(x)}(x^n - \overline{1}) \leq 1$ for all irreducible $q(x) \in \mathbb{Z}/p\mathbb{Z}[x]$.*

Proof. Working in $\mathbb{Z}/p\mathbb{Z}[x]$, we have $D(x^n - \overline{1}) = \overline{n} \cdot x^{n-1}$. Notice that $\overline{n} \neq \overline{0}$ because $p \nmid n$. Since $\mathbb{Z}/p\mathbb{Z}[x]$ is a Euclidean domain and $x \in \mathbb{Z}/p\mathbb{Z}[x]$ is irreducible, the only irreducible factors of $D(x^n - \overline{1})$ are the nonzero constant multiples of x (i.e. the associates of x). Since $x \nmid x^n - \overline{1}$ in $\mathbb{Z}/p\mathbb{Z}[x]$, it follows that $x^n - \overline{1}$ and $D(x^n - \overline{1})$ have no common irreducible factors in $\mathbb{Z}/p\mathbb{Z}[x]$, and hence are relatively prime in $\mathbb{Z}/p\mathbb{Z}[x]$. The result now follows from the previous proposition. □

Theorem 8.1.12. *$\Phi_n(x)$ is irreducible in $\mathbb{Q}[x]$ for all $n \in \mathbb{N}^+$.*

Proof. Let $g(x) \in \mathbb{Q}[x]$ be a monic irreducible factor of $\Phi_n(x)$ (such a factor exists because $\mathbb{Q}[x]$ is a Euclidean domain). Fix $h(x) \in \mathbb{Q}[x]$ with $\Phi_n(x) = g(x)h(x)$. By Lemma 8.1.7, we have $g(x), h(x) \in \mathbb{Z}[x]$ and also that $h(x)$ is monic. Fix an arbitrary root $u \in \mathbb{C}$ of $g(x)$, and notice that u is a primitive n^{th} root of unity because it is a root of $\Phi_n(x)$. Since $g(x)$ is irreducible in $\mathbb{Q}[x]$, it follows that $g(x)$ is the minimal polynomial of u over $\mathbb{Q}$.

Suppose that p is a prime with $p \nmid n$. We claim that u^p is also a root of $g(x)$. First notice that u^p is a primitive n^{th} root of unity (since $p \nmid n$ and hence $\gcd(p, n) = 1$). Thus, either $g(u^p) = 0$ or $h(u^p) = 0$. Assume that latter, i.e. assume that $h(u^p) = 0$. We then have that u is a root of $h(x^p)$, and since $g(x)$ is the minimal polynomial of u, we conclude that $g(x) \mid h(x^p)$ in $\mathbb{Q}[x]$. Fix $q(x) \in \mathbb{Q}[x]$ with

$$h(x^p) = g(x)q(x)$$

Now $h(x^p) \in \mathbb{Z}[x]$ and both $h(x^p)$ and $g(x)$ are monic, so $q(x) \in \mathbb{Z}[x]$ and is monic by Lemma 8.1.7. Reducing this equation modulo p, we see that

$$\overline{h}(x^p) = \overline{g}(x)\overline{q}(x)$$

in $\mathbb{Z}/p\mathbb{Z}[x]$. Now the ring $\mathbb{Z}/p\mathbb{Z}[x]$ has characteristic p and in $\mathbb{Z}/p\mathbb{Z}$ we have $a^p = a$ for all $a \in \mathbb{Z}/p\mathbb{Z}$ by Fermat's Little Theorem. Therefore, $\overline{h}(x^p) = (\overline{h}(x))^p$ by the Freshman's Dream Lemma. Hence in $\mathbb{Z}/p\mathbb{Z}[x]$ we have

$$(\overline{h}(x))^p = \overline{g}(x)\overline{q}(x)$$

Therefore, $\overline{g}(x) \mid (\overline{h}(x))^p$ in $\mathbb{Z}/p\mathbb{Z}[x]$, and hence $\overline{g}(x)$ and $\overline{h}(x)$ have a common irreducible factor in $\mathbb{Z}/p\mathbb{Z}[x]$ (since $\mathbb{Z}/p\mathbb{Z}[x]$ is a UFD). Now $\Phi_n(x) = g(x)h(x)$, hence $\overline{\Phi}_n(x) = \overline{g}(x)\overline{h}(x)$ in $\mathbb{Z}/p\mathbb{Z}[x]$, so it follows that $\overline{\Phi}_n(x)$ there exists an irreducible $\overline{m}(x) \in \mathbb{Z}/p\mathbb{Z}[x]$ such that $m(x)^2 \mid \overline{\Phi}_n(x)$ in $\mathbb{Z}/p\mathbb{Z}[x]$. Since $\overline{\Phi}_n(x) \mid x^n - \overline{1}$ in $\mathbb{Z}/p\mathbb{Z}[x]$, it follows that there exists an irreducible $\overline{m}(x) \in \mathbb{Z}/p\mathbb{Z}[x]$ such that $\overline{m}(x)^2 \mid x^n - \overline{1}$. This contradicts the previous corollary. Therefore, we can not have $h(u^p) = 0$, so we must have $g(u^p) = 0$.

Thus, for any root u of $g(x)$, and any prime $p \nmid n$, we have that u^p is also a root of $g(x)$. Now an arbitrary primitive n^{th} root of unity equals $\zeta_n^{p_1 p_2 \cdots p_k}$ for primes $p_i \nmid n$. Repeatedly applying the result, we conclude that that every primitive n^{th} root of unity is a root of $g(x)$. Since $g(x) \mid \Phi_n(x)$ in $\mathbb{Q}[x]$ and $g(x)$ is monic, this implies that $\Phi_n(x) = g(x)$. Therefore, $\Phi_n(x)$ is irreducible in $\mathbb{Q}[x]$. □

Corollary 8.1.13. *If $u \in \mathbb{C}$ be a primitive n^{th} root of unity, then $[\mathbb{Q}(u) : \mathbb{Q}] = \varphi(n)$. Thus, $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$.*

Lemma 8.1.14. *Let $n, k \in \mathbb{N}^+$. Suppose that $p \in \mathbb{N}^+$ is prime divisor of $\Phi_n(k)$. We then have*

- *$p \nmid k$.*
- *Either $p \mid n$ or $p \equiv 1 \pmod{n}$.*

Proof. Let $p \in \mathbb{N}^+$ be a prime divisor of $\Phi_n(k)$. Now in $\mathbb{Z}[x]$, we have

$$\Phi_n(x) \mid x^n - 1$$

so we can conclude that $\Phi_n(k) \mid k^n - 1$ in $\mathbb{Z}$. Since $p \mid \Phi_n(k)$, it follows that $p \mid (k^n - 1)$ in $\mathbb{Z}$, hence $k^n \equiv 1 \pmod{p}$. From this we can conclude that $p \nmid k$, which gives the first assertion.

Since $p \nmid k$, we have $\overline{k} \in U(\mathbb{Z}/p\mathbb{Z})$. Let m be the order of $\overline{k}$ viewed in the group $U(\mathbb{Z}/p\mathbb{Z})$. Working in the group $U(\mathbb{Z}/p\mathbb{Z})$, we have $\overline{k}^n = \overline{1}$, so $m \mid n$. We now have two cases.

- Suppose that $m = n$, i.e. the order of $\overline{k}$ in $U(\mathbb{Z}/p\mathbb{Z})$ is n. By Lagrange's Theorem, we conclude that $n \mid (p - 1)$, so $p \equiv 1 \pmod{n}$.

- Suppose that $m \neq n$, so m is proper divisor of n. We then have that $\overline{k}^m = \overline{1}$. Working in $\mathbb{Z}[x]$, there exists $g(x) \in \mathbb{Z}[x]$ with
$$x^n - 1 = \Phi_n(x) \cdot (x^m - 1) \cdot g(x)$$
(Here $g(x)$ is the product of all $\Phi_d(x)$ where d varies over all proper divisors of n that are not divisors of m; note that this might be an empty product in which case $g(x) = 1$). Reducing modulo p this gives
$$x^n - \overline{1} = \overline{\Phi}_n(x) \cdot (x^m - \overline{1}) \cdot \overline{g}(x)$$
Now $\overline{k}$ is a root of $x^m - \overline{1} \in \mathbb{Z}/p\mathbb{Z}[x]$ and also a root of $\overline{\Phi}_n(x) \in \mathbb{Z}/p\mathbb{Z}[x]$ (because $p \mid \Phi_n(k)$). This would imply that $(x - \overline{k})^2$ divides $x^n - \overline{1}$ in $\mathbb{Z}/p\mathbb{Z}[x]$. From the above corollary, this implies that $p \mid n$.

This completes the proof. □

Lemma 8.1.15. *Let $n \in \mathbb{N}$ with $n \geq 2$. We have $|\Phi_n(x)| > x - 1$ for all $x \in \mathbb{R}$ with $x \geq 2$.*

Proof. For any primitive n^{th} root of unity $u \in \mathbb{C}$, we have $|x - u| > x - 1$. Therefore
$$\begin{aligned} |\Phi_n(x)| &= \prod_{u \in P_n} |x - u| \\ &> \prod_{u \in P_n} (x - 1) \\ &= (x - 1)^{\varphi(n)} \\ &\geq x - 1 \end{aligned}$$
where the last line follows from the fact that $x - 1 \geq 1$ and $\varphi(n) \geq 1$. □

Theorem 8.1.16 (Dirichlet's Theorem on Primes in Arithmetic Progressions (Special Case))**.** *For every $n \in \mathbb{N}^+$, there exist infinitely many primes $p \in \mathbb{N}^+$ such that $p \equiv_n 1$.*

Proof. Let $n \in \mathbb{N}^+$. Fix an arbitrary $m \in \mathbb{N}$ with $m \geq 2$. We show that there is a prime $p > m$ with $p \equiv 1 \pmod{n}$. Notice that $mn \geq 2$, so by the previous lemma we know that $|\Phi_{mn}(mn)| \geq 2$. Fix a prime $p \in \mathbb{N}^+$ such that $p \mid \Phi_{mn}(mn)$. By the first part of the above lemma, we know that $p \nmid mn$. From this we conclude that $p \nmid n$, so using the second part of the above lemma, we conclude that $p \equiv 1 \pmod{mn}$. This implies both $p \equiv 1 \pmod{n}$ and also $p > mn \geq m$. □