Contents

8 Quadratic Integer Rings

Our goal in this chapter is describe various properties of quadratic integer rings, which are essentially the rings $\mathbb{Z}[\sqrt{D}]$ we have already encountered in our study of Pell's equation, along with some of their applications to number theory.

We begin with an overview of some properties of integral domains related to division algorithms, common divisors, and unique factorization; these topics are of independent number-theoretic interest since they will allow us to generalize many of the arithmetic properties of $\mathbb{Z}$. We then narrow our attention on the quadratic integer rings $\mathcal{O}_D$ with a goal of studying factorization in these rings. Although many of these rings do not have unique factorization of elements, we will prove that these rings do possess unique factorization of ideals (in the sense that every nonzero ideal is a unique product of prime ideals). We will then give some applications of these facts to classical problems in number theory.

8.1 Arithmetic in Rings and Domains

- In this section we will discuss some basic results from ring theory about Euclidean domains, ideals, and unique factorization.

8.1.1 Ideals of Commutative Rings

- We start by introducing ideals of commutative rings, which (in the study of general rings) are primarily motivated by their use in constructing quotient rings.
- Definition: If R is a commutative ring with 1, a subset I is called a (two-sided) ideal of R if it contains 0, is closed under subtraction, and is closed under arbitrary multiplication by elements of R. Explicitly, I is an ideal if I contains 0 and for any $x, y \in I$ and any $r \in R$, the elements $x - y$ and rx are in I.
  - We will mention that if R is a noncommutative ring, there are various other flavors of ideals (left ideals, right ideals, and two-sided ideals) that are not generally equivalent to one another. We will not deal with these since we are only interested in commutative rings.
  - There are various other ways to describe ideals. For example, I is an ideal of R if and only if I is a subgroup of R under addition that is also closed under arbitrary multiplication by elements of R.
- Here are a few basic examples of ideals:
  - Example: The subrings $n\mathbb{Z}$ are ideals of $\mathbb{Z}$, since they are clearly closed under arbitrary multiplication by elements of $\mathbb{Z}$.
  - Example: If $R = F[x]$ and p is any polynomial, the subring pR of multiples of p is an ideal of $F[x]$, since it is closed under arbitrary multiplication by polynomials in $F[x]$.
  - Non-example: The subring $\mathbb{Z}$ of $\mathbb{Q}$ is not an ideal of $\mathbb{Q}$, since it is not closed under arbitrary multiplication by elements of $\mathbb{Q}$. For example if we take $r = \frac{1}{3} \in \mathbb{Q}$ and $x = 4 \in \mathbb{Z}$, the element $rx = \frac{4}{3}$ is not in $\mathbb{Z}$.
  - Example: For any ring R, the subrings $\{0\}$ and R are ideals of R. We refer to $\{0\}$ as the trivial ideal (or the "zero ideal") and refer to any ideal $I \neq R$ as a proper ideal (since it is a proper subset of R).
- Here are a few more examples (and non-examples) of ideals.
- Example: In the polynomial ring $\mathbb{Z}[x]$, determine whether the set S of polynomials with even constant term (i.e., the polynomials of the form $2a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ for integers a_i) forms an ideal.
  - It is easy to see that $0 \in S$ and that S is closed under subtraction.
  - Furthermore, if $q(x)$ is any other polynomial, and $p(x) \in S$, then $p(x)q(x)$ also has even constant term, so it is also in S.
  - Thus, S is closed under multiplication by elements of $\mathbb{Z}[x]$, so it $\boxed{\text{is an ideal}}$.
- Example: Determine whether the set $S = \{\overline{0}, \overline{2}, \overline{4}, \overline{6}\}$ of "even" residue classes is an ideal of $\mathbb{Z}/8\mathbb{Z}$.
  - We have $0 \in S$, and it is a straightforward calculation to see that S is closed under subtraction, since the sum of two "even" residue classes modulo 8 will still be even.
  - Furthermore, the product of any residue class with an even residue class will again be an even residue class (since 8 is even), so S is closed under multiplication by arbitrary elements of R. Thus, S $\boxed{\text{is an ideal}}$.
- Example: Determine whether the set $S = \{(2a, 3a) : a \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z} \times \mathbb{Z}$.
  - We have $0 \in S$, and $(2a, 3a) - (2b, 3b) = (2(a - b), 3(a - b))$ so S is closed under subtraction.
  - But, for example, we can see that $(1, 2) \cdot (2, 3) = (2, 6)$ is not in S, even though $(2, 3)$ is, so S is not closed under arbitrary multiplication by elements of $\mathbb{Z} \times \mathbb{Z}$. Thus, S $\boxed{\text{is not an ideal}}$.
- In order to study the structure of ideals, we would like a simpler way to describe them. A convenient way is to describe ideals as being "generated" by subsets of a ring:

- If R is a ring with 1 and A is a subset of R, we would like to define "the ideal generated by A" to be the smallest ideal containing A.
- A priori, it is not obvious that there is such a smallest ideal. However, since the intersection of any nonempty collection of ideals is also an ideal, and since A is contained in at least one ideal (namely the whole ring R), we can equivalently define (A) to be the intersection of all ideals containing A.
- However, although the above analysis clearly indicates that these definitions are well-posed, we have not actually described what these ideals are.
- If I is the ideal generated by A, then if $a_1, a_2, \ldots, a_n$ are any elements of A, we see that I must contain the elements r_1a_1, r_2a_2, ... , r_na_n for any $r_i \in R$ and hence also contain their sum.
- On the other hand, if we let S be the set of elements of the form $r_1a_1 + r_2a_2 + \cdots + r_na_n$ for any $a_i \in A$ and $r_i \in R$ (and some $n \geq 0$), then it is easy to see that S is a subring that is closed under multiplication by elements of R, so S is an ideal.
- Furthermore, since R contains 1, S contains A, so S is an containing A hence must actually be the ideal generated by A.

- Our discussion above establishes the following proposition:

- Proposition (Generation of Ideals): Let R be a commutative ring with 1 and A be a subset of R. Then the set $(A) = \{r_1a_1 + r_2a_2 + \cdots + r_na_n : r_i \in R \text{ and } a_i \in A\}$ is the smallest ideal containing A.

- The simplest class of ideals are those generated by a finite set, and (in particular) those generated by a single element:

- Definition: If R is a ring with 1, we say an ideal I is finitely generated if I is generated by a finite set, and we say I is principal if I is generated by a single element. Thus, a finitely generated ideal has the form $I = (a_1, a_2, \ldots, a_n)$, while a principal ideal has the form $I = (a)$.
  - We emphasize here that the principal ideal (a) is simply the set of R-multiples of a: $(a) = \{ra : r \in R\}$.
  - Example: If R is any commutative ring with 1, then $R = (1)$ is principal. Likewise, the zero ideal $0 = (0)$ is also principal.
  - Example: In $\mathbb{Z}$, for any integer n we have $(n) = n\mathbb{Z}$. Since every ideal of $\mathbb{Z}$ is of the form $n\mathbb{Z}$, we see that every ideal of $\mathbb{Z}$ is principal. We remark that the notation $n\mathbb{Z}$ we have already used is consistent with the definition above.
  - Remark: If a and b are integers with greatest common divisor d, then $(a, b) = (d)$: this follows from the pair of observations that a and b are both contained in (d) so that $(a, b) \subseteq (d)$, and that $d = xa + yb$ for some integers x and y by the Euclidean algorithm, so that d is contained in (a, b). Indeed, as a reflection of this fact, many authors write (a, b) to denote the greatest common divisor of a and b.

- Since principal ideals are the easiest to describe, it is often useful to try to determine whether a particular ideal is principal, though this task is not always so easy! We give a few examples illustrating that this can often be a tricky question.

- Example: Show that the ideal $I = (2, x)$ in $\mathbb{Z}[x]$ is not principal.
  - Note that $I = \{2p(x) + xq(x) : p, q \in \mathbb{Z}[x]\}$ is the collection of polynomials in $\mathbb{Z}[x]$ with even constant term.
  - If I were principal and generated by some polynomial $r(x)$, then every polynomial in I would be divisible by $r(x)$. Hence, in particular, $r(x)$ would divide 2, so since 2 is a constant polynomial and a prime number, $r(x)$ would have to be one of $\{\pm 1, \pm 2\}$.
  - However, since $r(x)$ must also divide x, the only possibility is that $r(x)$ would be either 1 or -1. But it is easy to see that the ideal generated by 1 (or -1) is all of $\mathbb{Z}[x]$, so $r(x)$ cannot be 1 or -1, since $I \neq \mathbb{Z}[x]$.
  - Thus, there is no possible choice for r, so I is $\boxed{\text{not principal}}$. (Of course, it is still finitely generated!)

- Example: Determine whether or not the ideal $I = (2, 1 + \sqrt{-5})$ in $\mathbb{Z}[\sqrt{-5}]$ is principal.

- Suppose this ideal were principal with generator $r = a + b\sqrt{-5}$ in $\mathbb{Z}[\sqrt{-5}]$.
- Then r would necessarily divide 2, meaning that $2 = rs$ for some $s \in \mathbb{Z}[\sqrt{-5}]$. By taking norms, we see that $4 = N(2) = N(r)N(s)$.
- Likewise, since r divides $1+\sqrt{-5}$, we would have $1+\sqrt{-5} = rt$ for some $t \in \mathbb{Z}[\sqrt{-5}]$, so by taking norms we would have $6 = N(1+\sqrt{-5}) = N(r)N(t)$.
- Since $N(r) = a^2 + 5b^2$ is a nonnegative integer, we see that $N(r)$ must divide both 4 and 6, hence is either 1 or 2. However, it is easy to see that there are no integer solutions to $a^2 + 5b^2 = 2$, and the only elements of norm 1 are 1 and -1.
- As in the examples above, the ideal generated by 1 (or -1) is all of $\mathbb{Z}[\sqrt{-5}]$, but $(2, 1+\sqrt{-5}) \neq \mathbb{Z}[\sqrt{-5}]$ since every element $a + b\sqrt{-5}$ in the ideal has $a + b$ even.
- Thus, I is $\boxed{\text{not principal}}$.

- As we noted above, we always have $(1) = R$. We can in fact generalize this statement somewhat:
- Proposition (Ideals and Units): If I is an ideal of the ring R with 1, then $I = R$ if and only if I contains a unit.
  - Proof: If $I = R$ then certainly I contains a unit (namely, 1).
  - Conversely, if $u \in I$ is a unit with $ur = 1$, then since I is an ideal we have $1 = ur \in I$.
  - Then for any $s \in R$, the element $s = 1s$ is also in I, and so $I = R$.
- Since every nonzero element in a field is a unit, we immediately see that the only nonzero ideal of a field is the full ring. The converse is also true:
- Corollary (Ideals of Fields): A commutative ring R with 1 is a field if and only if the only ideals of R are 0 and R.
  - Proof: If F is a field and I is any nonzero ideal, then I contains some nonzero element r. Since F is a field, r is a unit, and so by the proposition above, $I = R$.
  - Conversely, if the only ideals of R are 0 and R, let $r \in R$ be any nonzero element. Then (r) contains $r \neq 0$ so it cannot be the zero ideal, so we must have $(r) = R$.
  - By the previous proposition, this means (r) contains 1: then $rs = 1$ for some $s \in R$, so r is a unit. Hence every nonzero element of R is a unit, so R is a field as claimed.

8.1.2 Quotient Rings

- Now that we have discussed ideals, we can use them to study residue classes, and thereby discuss construct quotient rings.
- Definition: If I is an ideal of the ring R, then we say a is congruent to b modulo I, written $a \equiv b \pmod{I}$, if $a - b \in I$.
  - As in $\mathbb{Z}$ and $F[x]$, congruence modulo I is an equivalence relation that respects addition and multiplication. The proofs are the same as in $\mathbb{Z}$ and $F[x]$, once we make the appropriate translations from "divisibility" to "containment in I".
- Proposition (Ideal Congruences): Let I be an ideal of R and $a, b, c, d \in R$. Then the following are true:
  1. $a \equiv a \pmod{I}$.
     - Proof: Since $a - a = 0 \in I$, the statement is immediate.
  2. $a \equiv b \pmod{I}$ if and only if $b \equiv a \pmod{I}$.
     - Proof: If $a - b \in I$ then $-(a-b) = b - a \in I$ since I is closed under additive inverses, and conversely if $b - a \in I$ then so is $-(b-a) = a - b$.
  3. If $a \equiv b \pmod{I}$ and $b \equiv c \pmod{I}$, then $a \equiv c \pmod{I}$.

- Proof: We are given $a-b \in I$ and $b-c \in I$, so since I is closed under addition, we see $(a-b)+(b-c) = a-c \in I$.

4. If $a \equiv b \pmod{I}$ and $c \equiv d \pmod{I}$, then $a+c \equiv b+d \pmod{I}$.
   - Proof: We are given $a-b \in I$ and $c-d \in I$, so since I is closed under addition, we see $(a-b)+(c-d) = (a+c)-(b+d) \in I$.
5. If $a \equiv b \pmod{I}$ and $c \equiv d \pmod{I}$, then $ac \equiv bd \pmod{I}$.
   - Proof: We are given $a-b \in I$ and $c-d \in I$. Then since I is closed under arbitrary left and right multiplication, we see that $(a-b)c$ and $b(c-d)$ are also in I. Hence $ac-bd = (a-b)c+b(c-d)$ is also in I since I is closed under addition.

- Now we can define residue classes:
- Definition: If I is an ideal of the ring R, then for any $a \in R$ we define the residue class of a modulo I to be the set $\overline{a} = a+I = \{a+x : x \in I\}$. This set is also called the coset of I represented by a.
  - We will use the notation $\overline{a}$ and $a+I$ interchangeably. (The latter is intended to evoke the idea of "adding" a to the set I.)
  - We observe, as with our previous examples of residue classes, that any two residue classes are either disjoint or identical and that they partition R: specifically, $\overline{a} = \overline{b}$ if and only if $a \equiv b \pmod{I}$ if and only if $a-b \in I$.
- All that remains is to verify that the residue classes form a ring, in the same way as in $\mathbb{Z}$ and $F[x]$:
- Theorem (Quotient Rings): Let I be an ideal of the ring R. Then the collection of residue classes modulo I forms a ring, denoted R/I (read as "R mod I"), under the operations $\overline{a}+\overline{b} = \overline{a+b}$ and $\overline{a} \cdot \overline{b} = \overline{ab}$. (This ring is called the quotient ring of R by I.) If R is commutative then so is R/I, and likewise if R has a 1 then so does R/I.
  - Remark: The notation R/I is intended to emphasize the idea that I represents a single element (namely, $\overline{0}$) in the quotient ring R/I, and the other elements in R/I are "translates" of I. In this way, R/I is the ring obtained from R by "collapsing" or "dividing out" by I, whence the name "quotient ring".
  - The proof of this fact is exactly the same as in the cases of $\mathbb{Z}$ and $F[x]$, and only requires showing that the operations are well-defined.
  - Proof: First we must show that the addition and multiplication operations are well-defined: that is, if we choose different elements $a' \in \bar{a}$ and $b' \in \bar{b}$, the residue class of $a'+b'$ is the same as that of $a+b$, and similarly for the product.
  - To see this, if $a' \in \bar{a}$ then $a' \equiv a \pmod{I}$, and similarly if $b' \in \bar{b}$ then $b' \equiv b \pmod{I}$.
  - Then $a'+b' \equiv a+b \pmod{I}$, so $\overline{a'+b'} = \overline{a+b}$. Likewise, $a'b' \equiv ab \pmod{I}$, so $\overline{a'b'} = \overline{ab}$.
  - Thus, the operations are well-defined.
  - For the ring axioms [R1]-[R6], we observe that associativity, commutativity, and the distributive laws follow immediately from the corresponding properties in R: the additive identity in R/I is $\bar{0}$ and the additive inverse of $\overline{a}$ is $\overline{-a}$.
  - Finally, if R is commutative then so will be the multiplication of the residue classes, and if R has a 1 then the residue class $\overline{1}$ is easily seen to be a multiplicative identity in R/I.
- This general description of "quotient rings" generalizes the two examples we have previously discussed: $\mathbb{Z}/m\mathbb{Z}$ and R/pR where $R = F[x]$.
  - To be explicit, $\mathbb{Z}/m\mathbb{Z}$ is the quotient of $\mathbb{Z}$ by the ideal $m\mathbb{Z}$, while $F[x]/p$ is the quotient of the polynomial ring $F[x]$ by the principal ideal (p) consisting of all multiples of p.
  - It is not hard to see that the integer congruence $a \equiv b \pmod{m}$, which we originally defined as being equivalent to the statement $m|(b-a)$, is the same as the congruence $a \equiv b \pmod{I}$ where I is the ideal $m\mathbb{Z}$, since $b-a \in m\mathbb{Z}$ precisely when $b-a$ is a multiple of m.

- Here are some additional examples of quotient rings:
- Example: If R is any ring, the quotient ring of R by the zero ideal, namely $R/0$, is (isomorphic to) R itself, while the quotient ring of R by itself, namely R/R, is (isomorphic to) the trivial ring $\{0\}$.
- Example: In $R = \mathbb{Z}[x]$, with I consisting of all multiples of $x^2 + 1$, describe the structure of the quotient ring R/I.
    - It is easy to see that I is an ideal of R, since it is a subring that is closed under arbitrary multiplication by elements of R.
    - From our discussion of polynomial rings, we know that the residue classes in R/I are represented uniquely by residue classes of the form $\overline{a + bx}$ where $a, b \in \mathbb{Z}$. Note that in this quotient ring, we have $\overline{x}^2 + \overline{1} = \overline{0}$, which is to say, $\overline{x}^2 = -\overline{1}$.
    - The addition in this quotient ring is given by $\overline{a + bx} + \overline{c + dx} = \overline{(a + c) + (b + d)x}$ while the multiplication is given by $\overline{a + bx} \cdot \overline{c + dx} = \overline{(ac - bd) + (ad + bc)x}$, which follows from the distributive law and the fact that $\overline{x}^2 = -\overline{1}$.
    - In this case, the quotient ring is isomorphic to the ring of Gaussian integers $\mathbb{Z}[i]$, with the isomorphism $\varphi : R/I \to \mathbb{Z}[i]$ given by $\varphi(\overline{a + bx}) = a + bi$.
- Example: In $R = \mathbb{Z}/8\mathbb{Z}$, with $I = \{0, 4\}$, describe the structure of the quotient ring R/I.
    - It is easy to see that I is an ideal of R, since it is a subring that is closed under arbitrary multiplication by elements of R. (Indeed, it is the principal ideal generated by 4.)
    - Since each residue class contains 2 elements, and R has 8 elements in total, there are four residue classes. With this observation in hand, it is not hard to give a list: $\overline{0} = I = \{0, 4\}$, $\overline{1} = 1 + I = \{1, 5\}$, $\overline{2} = 2 + I = \{2, 6\}$, and $\overline{3} = 3 + I = \{3, 7\}$.
    - Notice, for example, that in the quotient ring R/I, we have $\overline{1} + \overline{3} = \overline{0}$, $\overline{2} \cdot \overline{2} = \overline{0}$, and $\overline{2} \cdot \overline{3} = \overline{2}$: indeed, we can see that the structure of R/I is exactly the same as $\mathbb{Z}/4\mathbb{Z}$ (the labelings of the elements are even the same).
- We will also occasionally want to mention structure-preserving maps from one ring to another, which are called homomorphisms:
- Definition: A function $\varphi : R \to S$ is a ring homomorphism if $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$ and $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$ for all elements r_1 and r_2 in R. A homomorphism $\varphi : R \to S$ that is a bijection is called a ring isomorphism.
    - Example: The map $\varphi : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ defined by $\varphi(a) = \overline{a}$ is a ring homomorphism.
    - Example: If R is any ring, the map $\varphi : R \to R \times R$ given by $\varphi(r) = (r, r)$ is a ring homomorphism.
    - Example: The map $\varphi : \mathbb{Z}/6\mathbb{Z} \to (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ given by $\varphi(a) = (a \bmod 2, a \bmod 3)$ is a ring isomorphism.
- Associated to a homomorphism are two fundamental objects: the kernel and image.
- Definition: If $\varphi : R \to S$ is a ring homomorphism, the kernel of φ, denoted $\ker \varphi$, is the set of elements in R mapped to 0_S by φ. In other words, $\ker \varphi = \{r \in R : \varphi(r) = 0\}$.
    - Intuitively, the kernel measures how close φ is to being the zero map: if the kernel is large, then φ sends many elements to zero, while if the kernel is small, φ sends fewer elements to zero.
    - Example: The kernel of the reduction homomorphism $\varphi : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ with $\varphi(a) = \overline{a}$ is $m\mathbb{Z}$.
- Definition: If $\varphi : R \to S$ is a ring homomorphism, the image of φ, denoted $\operatorname{im} \varphi$, is the set of elements in S of the form $\varphi(r)$ for some $r \in R$.
    - In the context of general functions, the image is often called the range of φ.
    - Intuitively, the image measures how close φ is to being surjective: indeed (by definition) φ is surjective if and only if $\operatorname{im} \varphi = S$.

- One of the fundamental results about quotient rings is a relationship between homomorphisms and quotients:
- Theorem (First Isomorphism Theorem): If $\varphi : R \to S$ is a homomorphism of rings, then $R/\ker\varphi$ is isomorphic to $\operatorname{im}\varphi$.
    - Intuitively, φ is a surjective homomorphism $\varphi : R \to \operatorname{im}\varphi$. To turn it into an isomorphism, we must "collapse" its kernel to a single element: this is precisely what the quotient ring $R/\ker\varphi$ represents.
    - Proof: Let $I = \ker\varphi$. We use φ to construct a map $\psi : R/I \to \operatorname{im}\varphi$, and then show that it is injective and surjective.
    - The map is defined as follows: for any residue class $\overline{r} \in R/I$, we define $\psi(\overline{r}) = \varphi(r)$.
    - We must verify that this map ψ is well-defined, so suppose that r' is some other representative of the residue class $\overline{r}$: then $r' - r \in I$, so $\varphi(r' - r) = 0$ and thus $\varphi(r') = \varphi(r)$.
    - Thus, $\psi(\overline{r'}) = \varphi(r') = \varphi(r) = \psi(\overline{r})$, so the map ψ is well-defined.
    - It is then easy to see ψ is a homomorphism, since $\psi(\overline{r} + \overline{s}) = \varphi(r + s) = \varphi(r) + \varphi(s) = \psi(\overline{r}) + \psi(\overline{s})$ and likewise $\psi(\overline{r} \cdot \overline{s}) = \varphi(r \cdot s) = \varphi(r) \cdot \varphi(s) = \psi(\overline{r}) \cdot \psi(\overline{s})$.
    - Next, we see that $\psi(\overline{r}) = 0$ precisely when $\varphi(r) = 0$, which is to say $r \in \ker(\varphi)$, so that $\overline{r} = \overline{0}$. Thus, the only element in $\ker\psi$ is $\overline{0}$, so ψ is injective.
    - Finally, if s is any element of $\operatorname{im}\varphi$, then by definition there is some $r \in R$ with $\varphi(r) = s$: then $\psi(\overline{r}) = s$, meaning that ψ is surjective.
    - Since ψ is a homomorphism that is both injective and surjective, it is an isomorphism.

8.1.3 Maximal and Prime Ideals

- An important class of ideals are those that are "maximal" under inclusion (i.e., which are not contained in any other ideal except the full ring):
- Definition: If R is a ring, a maximal ideal of R is an ideal $M \neq R$ with the property that the only ideals of R containing M are M and R.
    - Example: If F is a field, then since the only ideals of F are 0 and F, the zero ideal is a maximal ideal of F.
    - Example: In $\mathbb{Z}$, the ideal $m\mathbb{Z}$ is contained in $n\mathbb{Z}$ precisely when n divides m. Accordingly, the maximal ideals of $\mathbb{Z}$ are precisely the ideals of the form $p\mathbb{Z}$, where p is a prime.
    - Non-example: The ideal (x) is not a maximal ideal of $\mathbb{Z}[x]$ because it is contained in the proper ideal $(2, x)$.
- A commutative ring with 1 must have maximal ideals:
- Theorem (Existence of Maximal Ideals): If R is a commutative ring with 1, then any proper ideal of R is contained in a maximal ideal.
    - Like a number of other general existence theorems (e.g., the proof that every vector space has a basis), this proof requires the (in)famous "axiom of choice" from set theory. The version of the axiom of choice typically used in algebra is known as Zorn's lemma: if S is a nonempty partially ordered set with the property that every chain in S has an upper bound, then S contains a maximal element[1].
    - Proof: Suppose R is a ring with 1 and I is a proper ideal of R.
    - Let S be the set of all proper ideals of R containing I, partially ordered under inclusion. Since $I \in S$, S is nonempty.
    - If C is any nonempty chain in S, let J be the union of all ideals in C: then $0 \in J$ since 0 is contained in any ideal in C.

[1] A partial ordering on a set S a relation $\leq$ such that for any $x, y, z \in S$, (i) $x \leq x$ (ii) $x \leq y$ and $y \leq x$ implies $x = y$, and (iii) $x \leq y$ and $y \leq z$ implies $x \leq z$. If S is a partially-ordered set, a subset C is a chain if for any $x, y \in C$, either $x \leq y$ or $y \leq x$, an upper bound for a subset B is an element $w \in B$ such that $b \leq w$ for all $b \in B$, and a maximal element of a subset B is an element $m \in B$ such that if $x \in B$ has $m \leq x$ then $m = x$.

- Furthermore, if $x, y \in J$ and $r \in R$, then by definition $x \in I_i$ and $y \in I_j$ for some I_i and I_j in C. Since $I_i \subseteq I_j$ or $I_j \subseteq I_i$ since C is a chain, it follows that $x - y$, rx, and xr are all in one of I_i or I_j, hence in J. Thus, J is an ideal.
- Also, if it were true that $J = R$, then the element 1 would be in J. But this is impossible, since by definition J is the union of a collection of proper ideals of R, none of which therefore contains 1.
- Therefore, J is an upper bound for S. Hence, by Zorn's lemma, J contains a maximal element, which is therefore a maximal ideal of R that contains I.

- It might initially appear to be difficult to detect whether a particular ideal is maximal. However, by using quotient rings, we can easily detect whether a given ideal is maximal:

- Proposition (Maximal Ideals and Quotients): If R is a commutative ring with 1, then the ideal M is maximal if and only if R/M is a field.

  - We will remark that this result is *not* true if we drop either of the assumptions on R (i.e., that it is commutative and has a 1).
  - Proof: It can be verified that there is a correspondence between ideals of R containing I and the ideals of R/I: if J is an ideal of R, then $\tilde{J} = \{j + I : j \in J\}$ is easily seen to be an ideal of R/I. Conversely, if we have any ideal $J/I = \{j + I : j \in J\}$ of R/I, it is straightforward to check that the collection of all elements $j \in R$ such that $j + I \in \tilde{J}$ is an ideal of R.
  - This means the ideals of R/M are in bijection with the ideals of R containing M: therefore, M is maximal precisely when the only ideals of R/M are 0 and R/M.
  - Furthermore, if R is commutative with 1, then R/M is also a commutative ring with 1, so R/M is a field if and only if the only ideals of R/M are 0 and R/M. Putting these two statements together yields the proposition.

- Corollary: If F is a field, the maximal ideals of $F[x]$ are precisely the principal ideals (p) where p is irreducible.

  - Proof: Every ideal of $F[x]$ is principal, and the quotient ring $F[x]/(p)$ is a field if and only if p is irreducible.

- Example: Determine whether the ideal $I = (2, x)$ is a maximal ideal of $R = \mathbb{Z}[x]$.

  - As we have already shown, the quotient ring $R/(2, x)$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, which is a field. Thus, I is a maximal ideal of R.

- Example: Determine whether the ideal $I = (2)$ is maximal in $R = \mathbb{Z}[\sqrt{2}]$.

  - In the quotient ring R/I, the residue class $\sqrt{2} + I$ is nonzero, but has the property that $(\sqrt{2} + I)^2 = 2 + I = 0 + I$ is equal to zero.
  - Thus, the quotient ring R/I has zero divisors hence is not a field, meaning that I is not a maximal ideal of R.

- In addition to maximal ideals, we have another important class of ideals in commutative rings:

- Definition: If R is a commutative ring with 1, a prime ideal of R is an ideal $P \neq R$ with the property that for any $a, b \in R$ with ab in P, at least one of a and b is in P.

  - As naturally suggested by the name, prime ideals are a generalization of the idea of a prime number in $\mathbb{Z}$: for $n > 1$, the ideal $n\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$ precisely when $ab \in n\mathbb{Z}$ implies $a \in n\mathbb{Z}$ or $b \in n\mathbb{Z}$. Equivalently (in the language of divisibility) this means $n|ab$ implies $n|a$ or $n|b$, and this is precisely the condition that n is either a prime number (or zero).
  - Example: The prime ideals of $\mathbb{Z}$ are (0) and the ideals $p\mathbb{Z}$ where p is a prime number.
  - A similar statement holds in $R = F[x]$: the ideal (p) is prime precisely when p is not a unit and $p|ab$ implies $p|a$ or $p|b$, and the latter condition is equivalent to saying that p is either irreducible or zero.
  - Example: The prime ideals of $F[x]$ are (0) and the ideals (p) where p is an irreducible polynomial of positive degree.

- Like with maximal ideals, there is an easy way to test whether an ideal is prime using quotient rings:
- Proposition (Prime Ideals and Quotients): If R is a commutative ring with 1, then the ideal P is prime if and only if R/P is an integral domain.
    - This proof is essentially just a restatement of the definition of a prime ideal using residue classes in the quotient ring using the observation that $r \in P$ if and only if $\overline{r} = \overline{0}$ in R/P.
    - Proof: If R is commutative with 1 and $P \neq R$, then R/P is also commutative with 1, so we need only test for zero divisors.
    - If P is a prime ideal, then $ab \in P$ implies $a \in P$ or $b \in P$. In the quotient ring, this says that $\overline{ab} = \overline{0}$ implies $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$, which is precisely the statement that R/P has no zero divisors.
    - Conversely, if R/P has no zero divisors, then $\overline{ab} = \overline{0}$ implies $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$, which is to say, $ab \in P$ implies $a \in P$ or $b \in P$. Furthermore, since R/P is not the zero ring (since this possibility is excluded by the definition of integral domain), we see $P \neq R$, and therefore P is a prime ideal of R.
- Corollary: A commutative ring with 1 is an integral domain if and only if 0 is a prime ideal.
    - Proof: 0 is prime if and only if the quotient $R/0 \cong R$ is an integral domain.
- Corollary: In a commutative ring with 1, every maximal ideal is prime.
    - Proof: If M is a maximal ideal, then R/M is a field. Every field is an integral domain, so M is a prime ideal.
- Example: Determine whether the ideals (x) and (x^2) in $\mathbb{Z}[x]$ are prime ideals.
    - By the division algorithm, the residue classes in $\mathbb{Z}[x]/(x)$ are of the form $\overline{a}$ for $a \in \mathbb{Z}$. Clearly, $\overline{a}+\overline{b} = \overline{a+b}$ and $\overline{a} \cdot \overline{b} = \overline{ab}$, so the arithmetic of the residue classes is the same as the arithmetic of $\mathbb{Z}$. This means $\mathbb{Z}[x]/(x)$ is an integral domain, so (x) is a prime ideal.
    - On the other hand, also by the division algorithm, we see that the residue classes in $\mathbb{Z}[x]/(x^2)$ are of the form $\overline{a+bx}$ where $a, b \in \mathbb{Z}$. Since $\overline{x} \cdot \overline{x} = \overline{0}$ but $\overline{x} \neq \overline{0}$, we see that $\mathbb{Z}[x]/(x^2)$ has zero divisors, and so (x^2) is not a prime ideal.

8.1.4 Arithmetic in Integral Domains

- We now discuss some properties of arithmetic in integral domains.
- Definition: Suppose that R is an integral domain and $a, b, d \in R$.
    1. We say that d divides a, written $d|a$, if there exists some $r \in R$ such that $a = rd$.
    2. We say d is a common divisor of a and b if $d|a$ and $d|b$.
    3. We say that a common divisor $d \in R$ is a greatest common divisor of a and b if $d \neq 0$ and for any other common divisor d', it is true that $d'|d$.
    4. If 1 is a greatest common divisor of a and b, then we say a and b are relatively prime.
    5. If $a = ub$ for some unit u, then we say a and b are associates.

    - Observe that every ring element divides each of its associates, and that "being associate" is an equivalence relation.
    - Two elements in an integral domain may not possess a greatest common divisor. If a and b do have a greatest common divisor d, then the collection of greatest common divisors of a and b is precisely the set of associates of d.
- Here is an explicit example of elements in an integral domain that do not possess a greatest common divisor:
- Example: Show that $2 + 2\sqrt{-5}$ and 6 do not possess a greatest common divisor in $\mathbb{Z}[\sqrt{-5}]$.
    - First, observe that 2 and $1 + \sqrt{-5}$ are both common divisors of $2 + 2\sqrt{-5}$ and 6.

- Now suppose that $2+2\sqrt{-5}$ and 6 had a gcd d: then d would divide $2(1+\sqrt{-5})$ and 6, and also be divisible by 2 and $1+\sqrt{-5}$.
- By taking norms, we see that $N(d)$ divides both $N(2+2\sqrt{-5}) = 24$ and $N(6) = 36$, hence divides 12.
- Also, $N(d)$ would also necessarily be a multiple of $N(2) = 4$ and $N(1+\sqrt{-5}) = 6$, hence be a multiple of 12.
- The only possibility is $N(d) = 12$, but there are no elements of norm 12 in $\mathbb{Z}[\sqrt{-5}]$, since there are no integer solutions to $a^2 + 5b^2 = 12$. This is a contradiction, so $2+2\sqrt{-5}$ and 6 do not possess a greatest common divisor in $\mathbb{Z}[\sqrt{-5}]$.

- Proposition (Properties of Divisibility): Let R be an integral domain. Then for any elements $a, b, d \in R$, the following are true:

1. The element d divides a if and only if the principal ideal (a) is contained in the principal ideal (d).
   - Proof: Note $(a) \subseteq (d)$ if and only if $a \in (d)$ if and only if $a = dk$ for some $k \in R$.
2. The elements a and b are associate if and only if $a|b$ and $b|a$, if and only if $(a) = (b)$.
   - Proof: Note $(a) = (b)$ if and only if $(a) \subseteq (b)$ and $(b) \subseteq (a)$, which is equivalent to $a|b$ and $b|a$ by (1). Furthermore, $a = ub$ for some unit u clearly implies $a|b$ and $b|a$. Conversely, if $a|b$ and $b|a$, then $a = rb$ and $b = sa$ for some r, s, and then $a = rsa$. If $a = 0$ then $b = 0$ also and we are done; otherwise we may cancel to see $rs = 1$ and so r is a unit.
3. If a and b have a gcd d, then the collection of greatest common divisors of a and b is precisely the set of associates of d.
   - Proof: If d is a gcd of a and b and u is any unit, then $(ud)|a$ and $(ud)|b$, and also if $d'|d$ then $d'|(ud)$ so ud is also a gcd. Furthermore, if d and e are both gcds of a and b, then $d|e$ and $e|d$ so that d and e are associates by (2).
4. The element d is a gcd of a and b if and only if (d) is the smallest principal ideal containing (a, b). In particular, if (a, b) is a principal ideal, then any generator is a gcd of a and b.
   - Proof: By (1) above, d is a common divisor of a and b if and only if (d) contains both (a) and (b), which is equivalent to saying $(a, b) \subseteq (d)$.
   - Then by (1) again, if d is a gcd of a and b and d' is any other common divisor, we must have $(d) \subseteq (d')$: thus, d is a gcd of a and b if and only if (d) is the smallest principal ideal containing (a, b).
   - Finally, if $(a, b) = (d)$ is itself principal, then clearly (d) is the smallest principal ideal containing (a, b).
   - Remark: The fact that $(a, b) = (d)$ if d is a gcd of a and b is the reason that the greatest common divisor is often denoted by the symbol (a, b).

- Now that we have established some basic properties of divisibility, we can talk about factorizations.

- Definition: Let R be an integral domain. A nonzero element $r \in R$ is irreducible if it is not a unit and, for any "factorization" $r = bc$ with $b, c \in R$, one of b and c must be a unit. A ring element that is not irreducible and not a unit is called reducible: it can be written as $r = ab$ where neither a nor b is a unit.

  - Example: The irreducible elements of $\mathbb{Z}$ are precisely the prime numbers (and their negatives).
  - Example: The irreducible elements of $F[x]$ are the irreducible polynomials of positive degree.
  - Example: The element 5 is reducible in $\mathbb{Z}[i]$, since we can write $5 = (2+i)(2-i)$ and neither $2+i$ nor $2-i$ is a unit in $\mathbb{Z}[i]$. However, the element $2+i$ is irreducible: if $2+i = bc$ for some $z, w \in \mathbb{Z}[i]$, then taking norms yields $5 = N(2+i) = N(b)N(c)$, and since 5 is a prime number, one of $N(b)$ and $N(c)$ would necessarily be ± 1, and then b or c would be a unit. Likewise, $2-i$ is also irreducible.
  - Example: The element 2 is irreducible in $\mathbb{Z}[\sqrt{-5}]$: if $2 = bc$ then taking norms yields $4 = N(2) = N(b)N(c)$, and since there are no elements of norm 2 in $\mathbb{Z}[\sqrt{-5}]$, one of $N(b)$ and $N(c)$ would necessarily be ± 1, and then b or c would be a unit.

- Inside $\mathbb{Z}$, the irreducible elements are the prime numbers. However, we have a different notion of a prime element in an arbitrary integral domain:

- Definition: Let R be an integral domain. A nonzero element $p \in R$ is prime if p is nonzero and not a unit, and for any $a, b \in R$, if $p|ab$ then $p|a$ or $p|b$. Equivalently, p is prime if p is nonzero and the ideal (p) is a prime ideal of R.
  - Example: The prime elements of $\mathbb{Z}$ are precisely the prime numbers (and their negatives).
  - Example: The prime elements of $F[x]$ are the irreducible polynomials of positive degree.
  - Example: The element $2+i$ is prime in $\mathbb{Z}[i]$: by the calculation above, if $ab \in (2+i)$ then $2+i = bc$ for some $z, w \in \mathbb{Z}[i]$, then taking norms yields $5 = N(2+i) = N(b)N(c)$, and since 5 is a prime number, one of $N(b)$ and $N(c)$ would necessarily be ± 1, and then b or c would be a unit.
  - Non-Example: The element 2 is not prime in $\mathbb{Z}[\sqrt{-5}]$: note that $6 = (1+\sqrt{-5})(1-\sqrt{-5})$ is divisible by 2, but neither $1+\sqrt{-5}$ nor $1-\sqrt{-5}$ is divisible by 2.
- As suggested by the examples above, prime elements are always irreducible, but irreducible elements are not necessarily prime (we will later discuss under what conditions irreducible elements will be prime):
- Proposition (Primes are Irreducible): In an integral domain, prime elements are always irreducible.
  - Proof: Suppose $p \in R$ is a prime element. If $p = bc$ then since $p|bc$, we conclude that $p|b$ or $p|c$; without loss of generality suppose $b = pr$.
  - Then $p = prc$, so since $p \neq 0$ we may cancel to conclude $rc = 1$, so that c is a unit. Thus, p is irreducible.

8.1.5 Quadratic Fields and Quadratic Integer Rings

- We can now discuss some facts about the rings that we will be analyzing in this chapter. First, we need to mention quadratic fields:
- Definition: Let D be a squarefree integer not equal to 1. The quadratic field $\mathbb{Q}(\sqrt{D})$ is the set of complex numbers of the form $a + b\sqrt{D}$, where a and b are rational numbers.
  - Remark: An integer is squarefree if it is not divisible by the square of any prime, and not equal to 1. We lose nothing here by assuming that D is a squarefree integer, since two different integers differing by a square factor would generate the same set of complex numbers $a + b\sqrt{D}$.
  - The arithmetic in $\mathbb{Q}(\sqrt{D})$ is as follows: $(a+b\sqrt{D})+(c+d\sqrt{D}) = (a+c)+(b+d)\sqrt{D}$, and $(a+b\sqrt{D})(c+d\sqrt{D}) = (ac+Dbd)+(ad+bc)\sqrt{D}$.
  - Since $\mathbb{Q}(\sqrt{D})$ is clearly closed under subtraction and multiplication, and contains $0 = 0 + 0\sqrt{D}$, it is a subring of $\mathbb{C}$ and hence an integral domain, since it contains 1.
  - It is in fact a field (justifying the name "quadratic field") because we can write $(a+b\sqrt{D})^{-1} = \dfrac{a-b\sqrt{D}}{a^2-Db^2}$, and $a^2 - Db^2 \neq 0$ provided that a and b are not both zero because $\sqrt{D}$ is irrational by the assumption that D is squarefree and not equal to 1.
  - We will also remark that $\mathbb{Q}(\sqrt{D})$ is isomorphic to the quotient ring $\mathbb{Q}[x]$ modulo the principal ideal $(x^2 - D)$, with the isomorphism given explicitly by mapping $p(x) \in \mathbb{Q}[x]$ to $p(\sqrt{D}) \in \mathbb{Q}(\sqrt{D})$.
- Definition: The field norm $N : \mathbb{Q}(\sqrt{D}) \to \mathbb{Q}$ is defined to be the function $N(a + b\sqrt{D}) = a^2 - Db^2 = (a + b\sqrt{D})(a - b\sqrt{D})$.
  - The fundamental property of this field norm is that it is multiplicative: $N(xy) = N(x)N(y)$ for two elements x and y in $\mathbb{Q}(\sqrt{D})$, as can be verified by writing out both sides explicitly and comparing the results.
  - The field norm provides a measure of "size" of an element of $\mathbb{Q}(\sqrt{D})$, in much the same way that the complex absolute value measures the "size" of a complex number. In fact, if $D < 0$, then the field norm of an element $a + b\sqrt{D}$ is the same as the square of its complex absolute value.
- A fundamental subring of the quadratic field $\mathbb{Q}(\sqrt{D})$ is its associated "quadratic integer ring".

- The most obvious choice for an analogy of the integers $\mathbb{Z}$ inside $\mathbb{Q}(\sqrt{D})$ would be the set $\mathbb{Z}[\sqrt{D}] = \{a + b\sqrt{D} : a, b \in \mathbb{Z}\}$.
- However, notice that if $D \equiv 1 \pmod 4$, then the slightly larger subset $\mathbb{Z}[\frac{1+\sqrt{D}}{2}] = \{a + b\frac{1+\sqrt{D}}{2} : a, b \in \mathbb{Z}\}$ is actually also a subring: closure under subtraction is obvious, and for multiplication we can write $(a + b\frac{1+\sqrt{D}}{2})(c + d\frac{1+\sqrt{D}}{2}) = (ac + \frac{D-1}{4}bd) + (ad + bc + bd)\frac{1+\sqrt{D}}{2}$.
- One reason that this slightly larger set turns out to give a better analogy for the integers $\mathbb{Z}$ when $D \equiv 1 \pmod 4$ is that the number $\frac{1+\sqrt{D}}{2}$ satisfies a polynomial with integer coefficients and leading coefficient 1: explicitly, it is a root of $x^2 - x + \frac{1-D}{4} = 0$.

- Definition: The ring of integers $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ in the quadratic field $\mathbb{Q}(\sqrt{D})$ is defined as $\mathbb{Z}[\sqrt{D}]$ if $D \equiv 2$ or $3 \pmod 4$ and as $\mathbb{Z}[\frac{1+\sqrt{D}}{2}]$ if $D \equiv 1 \pmod 4$. Each of these rings is an integral domain.
  - For $D \equiv 2, 3 \pmod 4$, observe that $N(a + b\sqrt{D}) = a^2 - Db^2$ is an integer for every $a + b\sqrt{D} \in \mathcal{O}_{\sqrt{D}}$.
  - Likewise, if $D \equiv 1 \pmod 4$, we have $N(a + b\frac{1+\sqrt{D}}{2}) = a^2 + ab + \frac{1-D}{4}b^2$ is also an integer for every $a + b\frac{1+\sqrt{D}}{2} \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$.
  - Thus, the field norm N is always integer-valued on $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$. We can in fact use it to determine whether a given element is a unit:

- The units in the quadratic integer rings are the elements of norm ± 1:

- Proposition (Characterizing Units in $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$): An element r in the ring $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is a unit if and only if $N(r) = \pm 1$.
  - Proof: Suppose $r = a + b\sqrt{D}$ and let $\overline{r} = a - b\sqrt{D}$, so that $N(r) = r\overline{r}$. (Note that $\overline{r} = 2a - r$, so that even when $D \equiv 1 \pmod 4$, so that a and b are possibly half-integers, we see that $\overline{r}$ is still in $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$.)
  - If $N(r) = \pm 1$, then we see that $r\overline{r} = \pm 1$, so (by multiplying by -1 if necessary) we obtain a multiplicative inverse for r.
  - Conversely, suppose r is a unit and $rs = 1$. Taking norms yields $N(r)N(s) = N(rs) = 1$. Since $N(r)$ and $N(s)$ are both integers, we see that $N(r)$ must either be 1 or -1.

- Example: Find the units in $\mathbb{Z}[i]$ and $\mathbb{Z}[(1 + \sqrt{-3})/2]$.
  - For $\mathbb{Z}[i]$, we have $D = -1$, so if $r = a + bi$ we see $N(r) = a^2 + b^2$. We must therefore solve $a^2 + b^2 = 1$ in $\mathbb{Z}$: there are clearly four solutions, corresponding to $r = \boxed{1,\ i,\ -1,\ -i}$.
  - For $\mathbb{Z}[(1 + \sqrt{-3})/2]$, we have $D = -3$, so if $r = a + b\frac{1+\sqrt{-3}}{2}$ we see $N(r) = a^2 + ab + b^2$. We must therefore solve $a^2 + ab + b^2 = 1$ in $\mathbb{Z}$: by multiplying by 4 and completing the square, this equation is equivalent to $(2a + b)^2 + 3b^2 = 4$, which has six solutions corresponding to $r = \boxed{1,\ -1,\ \omega,\ -\omega,\ \omega^2,\ -\omega^2}$, where $\omega = \frac{1+\sqrt{-3}}{2}$ is seen to be a sixth root of unity satisfying $\omega^6 = 1$.

- In general, determining the full set of units in $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is a nontrivial computation that essentially reduces to solving Pell's equation.
  - When $D < 0$ it is not too difficult to see (by completing the square in a similar way to above when $D \equiv 1 \pmod 4$) that if $D \neq -1, -3$, then the only units in $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ are ± 1.
  - When $D > 0$ and $D \equiv 2, 3 \pmod 4$, solving $N(\alpha) = \pm 1$ is equivalent to solving Pell's equation $x^2 - Dy^2 = \pm 1$, which we have already described at length.

- For $D > 0$ with $D \equiv 1 \pmod 4$, we see $N(\frac{a+b\sqrt{D}}{2}) = \pm 1$ is equivalent to the Pell's equation $a^2 - Db^2 = \pm 4$, whose solutions (per our analysis) can also be found using continued fractions.
- In particular, the same sort of analysis we gave for $x^2 - Dy^2 = \pm 1$ will show that the solutions are of the form $\pm u^n$ where u is the fundamental unit.

- By using norms, we can also study possible factorizations and establish the irreducibility of elements. The following special case is often helpful:

- Proposition (Some Irreducibles in $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$): If $r \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ has $N(r) = \pm p$ where p is a prime number, then r is irreducible in $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$.

  - Proof: Suppose $N(r) = \pm p$ and we had a factorization $r = s_1 s_2$. Taking norms yields $\pm p = N(s_1 s_2) = N(s_1)N(s_2)$.
  - But since p is prime and $N(s_1)$ and $N(s_2)$ are integers, the only possibility is to have one of $N(s_1)$ and $N(s_2)$ equal to ± 1, which by our result earlier means that s_1 or s_2 is a unit. Then r is indeed irreducible, as claimed.

- Here are some examples of how we can establish irreducibility by computing norms:

  - Example: The elements $1 + i$ and $2 + i$ in $\mathbb{Z}[i]$ are irreducible, since their norms are 2 and 5 respectively.
  - Example: The elements $\frac{5+\sqrt{5}}{2}$ and $4 + \sqrt{5}$ in $\mathcal{O}_{\mathbb{Q}(\sqrt{5})}$ are irreducible since their norms are 5 and 11 respectively.
  - We remark that the proposition is not an if-and-only-if, as there can exist irreducible elements of non-prime norm as well.
  - Example: The element $3 \in \mathbb{Z}[i]$ has $N(3) = 9$, but 3 is irreducible because any factorization $3 = z_1 z_2$ would require $9 = N(3) = N(z_1)N(z_2)$, but since there are no elements of norm 3 in $\mathbb{Z}[i]$, the only possible factorizations require $N(z_1)$ or $N(z_2)$ to equal 1.
  - Example: The element $1 + \sqrt{-5} \in \mathcal{O}_{\mathbb{Q}(\sqrt{-5})}$ has $N(1 + \sqrt{-5}) = 6$, but $1 + \sqrt{-5}$ is irreducible because any factorization would have to be into a product of an element of norm 2 and an element of norm 3, but there are no such elements in $\mathcal{O}_{\mathbb{Q}(\sqrt{-5})}$.

- We will discuss more about factorization in these rings after we have developed some additional results about ideals and factorizations in general rings. For notational convenience, we will often write $\mathcal{O}_{\sqrt{D}}$ as shorthand for $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$.

8.1.6 Euclidean Domains

- Our next goal is to discuss what it means for an integral domain to possess a "division algorithm":

- Definition: If R is an integral domain, any function $N : R \to \{0, 1, 2, \dots\}$ such that $N(0) = 0$ is called a norm on R.

  - Observe that this is a rather weak property, and that any given domain may possess many different norms.

- Definition: A Euclidean domain (or domain with a division algorithm) is an integral domain R that possesses a norm N with the property that, for every a and b in R with $b \neq 0$, there exist some q and r in R such that $a = qb + r$ and either $r = 0$ or $N(r) < N(b)$.

  - The purpose of the norm function is to allow us to compare the size of the remainder to the size of the original element. Note that the quotient and remainder are *not* required to be unique!
  - Example: Any field is a Euclidean domain, because any norm will satisfy the defining condition. This follows because for every a and b with $b \neq 0$, we can write $a = qb + 0$ with $q = a \cdot b^{-1}$.
  - Example: The integers $\mathbb{Z}$ are a Euclidean domain with $N(n) = |n|$.

- Example: If F is a field, then the polynomial ring $F[x]$ is a Euclidean domain with norm given by $N(p) = \deg(p)$ for $p \neq 0$.

- Before we give additional examples, we will remark that the reason Euclidean domains have that name is that we can perform the Euclidean algorithm in such a ring, in precisely the same manner as in $\mathbb{Z}$ and $F[x]$:

- Definition: If R is a Euclidean domain, then for any $a, b \in R$ with $b \neq 0$, the Euclidean algorithm in R consists of repeatedly applying the division algorithm to a and b as follows, until a remainder of zero is obtained:

$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{k-1} &= q_k r_k + r_{k+1} \\ r_k &= q_{k+1} r_{k+1}. \end{aligned}$$

  - By the construction of the division algorithm, we know that $N(r_1) > N(r_2) > \cdots$, and since $N(r_i)$ is a nonnegative integer for each i, this sequence must eventually terminate with the last remainder equalling zero (else we would have an infinite decreasing sequence of nonnegative integers).

- The Gaussian integers provide another important example of a Euclidean domain:

- Proposition ($\mathbb{Z}[i]$ is Euclidean): The Gaussian integers $\mathbb{Z}[i]$ are a Euclidean domain, under the norm $N(a+bi) = a^2 + b^2$.

  - Explicitly, given $a + bi$ and $c + di$ in $\mathbb{Z}[i]$, we will describe how to produce $q, r \in \mathbb{Z}[i]$ such that $a + bi = q(c + di) + r$, and $N(r) \leq \frac{1}{2}N(c + di)$. This is even stronger than is needed (once we note that the only element of norm 0 is 0).
  - Proof: We need to describe the algorithm for producing q and r when dividing an element $a + bi$ by an element $c + di$.
  - If $c+di \neq 0$, then we can write $\dfrac{a+bi}{c+di} = x+iy$ where $x = (ac+bd)/(c^2+d^2)$ and $y = (bc-ad)/(c^2+d^2)$ are real numbers.
  - Now we define $q = s + ti$ where s is the integer closest to x and t is the integer closest to y, and set $r = (a + bi) - q(c + di)$. Clearly, $(a + bi) = q(c + di) + r$.
  - All we need to do now is show $N(r) \leq \frac{1}{2}N(c+di)$: first observe that $\dfrac{r}{c+di} = \dfrac{a+bi}{c+di} - q = (x-s)+(y-t)i$. Then because $|x - s| \leq \frac{1}{2}$ and $|y - t| \leq \frac{1}{2}$ by construction, the triangle inequality implies $\left|\dfrac{r}{c+di}\right| \leq \dfrac{\sqrt{2}}{2}$. Squaring both sides and rearranging yields $N(r) \leq \frac{1}{2}N(c + di)$, as desired.
  - Remark: For other quadratic integer rings $\mathcal{O}_{\sqrt{D}}$, the function $N(a + b\sqrt{D}) = \left|a^2 - Db^2\right|$ is a norm, but it does not in general give a division algorithm. The proof given above can, however, be adapted fairly easily to show that $\mathcal{O}_{\sqrt{D}}$ is a Euclidean domain for certain other small values of D, such as $D = -7$, -3, -2, and 2.

- As in $\mathbb{Z}$ and $F[x]$, we may also use the Euclidean algorithm to compute gcds:

- Theorem (Bézout): If R is a Euclidean domain and a and b are arbitrary elements with $b \neq 0$, then the last nonzero remainder d arising from the Euclidean Algorithm applied to a and b is a greatest common divisor of a and b. (In particular, any two elements in a Euclidean domain always possess at least one gcd.) Furthermore, there exist elements $x, y \in R$ such that $d = ax + by$.

  - The ideas in the proof are the same as for the proofs over $\mathbb{Z}$ and $F[x]$.
  - Proof: By an easy induction (starting with $r_k = q_{k+1}r_{k+1}$), $d = r_{k+1}$ divides r_i for each $1 \leq i \leq k$. Thus we see $d|a$ and $d|b$, so the last nonzero remainder is a common divisor.

- Suppose d' is some other common divisor of a and b. By another easy induction (starting with $d'|(a - q_1 b) = r_1$), it is easy to see that d' divides r_i for each $1 \leq i \leq k+1$, and therefore $d'|d$. Hence d is a greatest common divisor.
- For the existence of x and y with $d = ax + by$, we simply observe (by yet another easy induction starting with $r_1 = a - q_1 b$) that each remainder can be written in the form $r_i = x_i a + y_i b$ for some $x_i, y_i \in R$.

- Example: Find a greatest common divisor of $50 - 50i$ and $43 - i$ in $\mathbb{Z}[i]$, and write it in the form $d = (50 - 50i)x + (43 - i)y$ for some $x, y \in \mathbb{Z}[i]$.
  - We use the Euclidean algorithm. Dividing $43 - i$ into $50 - 50i$ yields $\dfrac{50 - 50i}{43 - i} = \dfrac{44}{37} - \dfrac{42}{37}i$, so rounding to the nearest Gaussian integer yields the quotient $q = 1 - i$. The remainder is then $50 - 50i - (1 - i)(43 - i) = (8 - 6i)$.
  - Next, dividing $8 - 6i$ into $43 - i$ yields $\dfrac{43 - i}{8 - 6i} = \dfrac{7}{2} + \dfrac{5}{2}i$, so rounding to the nearest Gaussian integer (there are four possibilities so we just choose one) yields the quotient $q = 3 + 2i$. The remainder is then $43 - i - (3 + 2i)(8 - 6i) = (7 + i)$.
  - Finally, dividing $7 + i$ into $8 - 6i$ yields $\dfrac{8 - 6i}{7 + i} = 1 - i$, so the quotient is $1 - i$ and the remainder is 0.
  - The last nonzero remainder is $\boxed{7 + i}$ so it is a gcd. To express the gcd as a linear combination, we solve for the remainders:

$$\begin{aligned} 8 - 6i &= 1 \cdot (50 - 50i) - (1 - i) \cdot (43 - i) \\ 7 + i &= (43 - i) - (3 + 2i)(8 - 6i) \\ &= (43 - i) - (3 + 2i) \cdot (50 - 50i) + (3 + 2i)(1 - i) \cdot (43 - i) \\ &= (-3 - 2i) \cdot (50 - 50i) + (6 - i) \cdot (43 - i) \end{aligned}$$

  and so we have $7 + i = \boxed{(-3 - 2i) \cdot (50 - 50i) + (6 - i) \cdot (43 - i)}$.

- The ideals of Euclidean domains are particularly simple:
- Theorem (Ideals of Euclidean Domains): Every ideal of a Euclidean domain is principal.
  - Proof: Clearly the zero ideal is principal, so suppose I is a nonzero ideal of the Euclidean domain R and let d be a nonzero element of I of smallest possible norm. (Such an element must exist by the well-ordering axiom.)
  - Since $d \in I$ we have $(d) \subseteq I$. If $a \in I$ is any other element, by the division algorithm we can write $a = qd + r$ for some r where either $r = 0$ or $N(r) < N(d)$.
  - However, since $r = a - qd \in I$ since both a and qd are in I, and $N(d)$ is minimal, we must have $r = 0$. Therefore, $a = qd$ and thus $a \in (d)$, so $I \subseteq (d)$. Hence $I = (d)$ is principal, as claimed.

- Corollary: Every ideal of $\mathbb{Z}$, $F[x]$, and $\mathbb{Z}[i]$ is principal, for any field F.
  - Proof: Each of these rings is a Euclidean domain.

- By the result above, we can deduce that any ring containing a non-principal ideal is not Euclidean (with respect to any norm):
  - Example: The ring $\mathbb{Z}[x]$ is not a Euclidean domain, since the ideal $(2, x)$ is not principal.
  - Example: The ring $\mathbb{Z}[\sqrt{-5}]$ is not a Euclidean domain, since the ideal $(2, 1 + \sqrt{-5})$ is not principal.

8.1.7 Principal Ideal Domains

- We have seen that every ideal in a Euclidean domain is principal. We now expand our attention to the more general class of rings in which every ideal is principal.
- Definition: A principal ideal domain (PID) is an integral domain in which every ideal is principal.
    - Example: As we have shown, every Euclidean domain is a principal ideal domain, so $\mathbb{Z}$, $\mathbb{Z}[i]$, and $F[x]$ are principal ideal domains.
    - Non-Example: The ring $\mathbb{Z}[x]$ is not a principal ideal domain, since the ideal $(2, x)$ is not principal.
    - Non-Example: The ring $\mathbb{Z}[\sqrt{-5}]$ is not a principal ideal domain, since the ideal $(2, 1 + \sqrt{-5})$ is not principal.
    - There exist principal ideal domains that are not Euclidean domains (although this is not so easy to prove). One example is the quadratic ring $\mathcal{O}_{\sqrt{-19}} = \mathbb{Z}[(1 + \sqrt{-19})/2]$.
- Like in Euclidean domains, we can show that any two elements in a PID have a greatest common divisor.
    - The substantial advantage of a Euclidean domain over a general PID is that we have an algorithm for computing greatest common divisors in Euclidean domains, rather than merely knowing that they exist.
- Proposition (Divisibility in PIDs): If R is a principal ideal domain and $a, b \in R$ are nonzero, then any generator d of the principal ideal (a, b) is a greatest common divisor of a and b. (In particular, any two elements in a principal ideal domain always possess at least one gcd.) Furthermore, there exist elements $x, y \in R$ such that $d = ax + by$.
    - Proof: We showed already that if (a, b) is principal, then any generator is a gcd of a and b. Furthermore, if $(a, b) = (d)$ then $d \in (a, b)$ implies that $d = ax + by$ for some $x, y \in R$.
- Our ultimate goal is to show that these rings (like the prototypical examples $\mathbb{Z}$ and $F[x]$) have the property that every nonzero element can be written as a finite product of irreducible elements, up to associates and reordering.
    - To show this, we will use essentially the same argument as in $\mathbb{Z}$ and $F[x]$: first we will prove that every element can be factored into a product of irreducibles, and then we will prove that the factorization is unique.
    - For the existence, if r is a reducible element then we can write $r = r_1 r_2$ where neither r_1 nor r_2 is a unit. If both r_1 and r_2 are irreducible, we are done: otherwise, we can continue factoring (say) $r_1 = r_{1,1} r_{1,2}$ with neither term a unit. If $r_{1,1}$ and $r_{1,2}$ are both irreducible, we are done: otherwise, we factor again.
    - We need to ensure that this process will always terminate: if not, we would obtain an infinite ascending chain of ideals $(r) \subset (r_1) \subset (r_{1,1}) \subset \cdots$, so first we will prove that this cannot occur.
    - Then to establish uniqueness, we use the same argument as in $\mathbb{Z}$ and $F[x]$: this requires showing that if p is irreducible, then $p|ab$ implies $p|a$ or $p|b$: in other words, that p is prime.
- First we establish the necessary result about ascending chains of ideals:
- Theorem (Ascending Chains in PIDs): If R is a principal ideal domain and the ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n \subseteq \cdots$ form an ascending chain, then there exists some positive integer N after which the chain is stationary: $I_n = I_N$ for all $n \geq N$.
    - Remark: A ring satisfying this "ascending chain condition" is called Noetherian.
    - Proof: Let J be the union of the ideals in the chain. We have shown already (in the course of proving that a ring with 1 always possesses maximal ideals) that the union of an ascending chain of ideals is also an ideal, so J is an ideal.
    - Since R is a PID, we see $J = (a)$ for some $a \in R$. But since J is a union, this means $a \in I_N$ for some N. But then for each $n \geq N$ we see $(a) = I_N \subseteq I_n \subseteq J = (a)$: we must have equality everywhere, so $I_n = I_N$ for all $n \geq N$.

- Next, we show that irreducible elements are prime:
- Proposition (Irreducibles are Prime in a PID): Every irreducible element in a principal ideal domain is prime.
  - Proof: Suppose that p is an irreducible element of R: to show that p is prime, we may equivalently show that the ideal (p) is a prime ideal.
  - So suppose (a) is an ideal containing (p): then $p \in (a)$ so $p = ra$ for some $r \in R$. But since p is irreducible, we either have $p|r$ or $p|a$, which is to say, either $r \in (p)$ or $a \in (p)$.
  - If $a \in (p)$ then $(a) \subseteq (p)$ and so $(a) = (p)$. Otherwise, if $r \in (p)$ then $r = sp$ for some $s \in R$, and then $p = ra$ implies $p = spa$, so since $p \neq 0$ we see $sa = 1$ and therefore a is a unit, and so $(a) = R$.
  - Thus, (a) is either (p) or R, meaning that (p) is a maximal hence prime ideal.
- In the proposition above, notice that we actually established that the prime element p generated a maximal ideal. This argument in fact shows that nonzero prime ideals are maximal in PIDs:
- Proposition (Prime Implies Maximal in a PID): Every nonzero prime ideal in a principal ideal domain is maximal.
  - Proof: Suppose that $I = (p)$ is a nonzero prime ideal of R, and suppose that (a) is an ideal containing I.
  - Since $p \in (a)$, we see that $p = ra$ for some $r \in R$. But then $ra \in (p)$, so since (p) is a prime ideal we either have $r \in (p)$ or $a \in (p)$.
  - By the same argument as in the proposition above, we conclude that (a) is either (p) or R, meaning that (p) is a maximal ideal.
- Now we can establish that principal ideal domains have unique factorization:
- Theorem (Unique Factorization in PIDs): If R is a principal ideal domain, then every nonzero nonunit $r \in R$ can be written as a finite product of irreducible elements. Furthermore, this factorization is unique up to associates: if $r = p_1 p_2 \cdots p_d = q_1 q_2 \cdots q_k$ for irreducibles p_i and q_j, then $d = k$ and there is some reordering of the factors such that p_i is associate to q_i for each $1 \leq i \leq k$.
  - Proof: Suppose $r \in R$ is not zero and not a unit.
  - If r is irreducible, we already have the required factorization. Otherwise, $r = r_1 r_2$ for some nonunits r_1 and r_2. If both r_1 and r_2 are irreducible, we are done: otherwise, we can continue factoring (say) $r_1 = r_{1,1} r_{1,2}$ with neither term a unit. If $r_{1,1}$ and $r_{1,2}$ are both irreducible, we are done: otherwise, we factor again.
  - We claim that this process must terminate eventually: otherwise (as follows by the axiom of choice), we would have an infinite chain of elements x_1, x_2, x_3, ... , such that $x_1|r$, $x_2|x_1$, $x_3|x_2$, and so forth, where no two elements are associates, yielding an infinite chain of ideals $(r) \subset (x_1) \subset (x_2) \subset \cdots$ with each ideal properly contained in the next. But this is impossible, since every ascending chain of ideals in R must become stationary.
  - Thus, the factoring process must terminate, and so r can be written as a product of irreducibles.
  - We establish uniqueness by induction on the number of irreducible factors of $r = p_1 p_2 \cdots p_n$.
  - If $n = 1$, then r is irreducible. If r had some other nontrivial factorization $r = qc$ with q irreducible, then q would divide r hence be associate to r (since irreducibles are prime). But this would mean that c is a unit, which is impossible.
  - Now suppose $n \geq 2$ and that $r = p_1 p_2 \cdots p_d = q_1 q_2 \cdots q_k$ has two factorizations into irreducibles.
  - Since $p_1|(q_1 \cdots q_k)$ and p_1 is irreducible hence prime, repeatedly applying the fact that p irreducible and $p|ab$ implies $p|a$ or $p|b$ shows that p_1 must divide q_i for some i.
  - By rearranging we may assume $q_1 = p_1 u$ for some u: then since q_1 is irreducible (and p_1 is not a unit), u must be a unit, so p_1 and q_1 are associates.
  - Cancelling then yields the equation $p_2 \cdots p_d = (uq_2) \cdots q_k$, which is a product of fewer irreducibles. By the induction hypothesis, such a factorization is unique up to associates. This immediately yields the desired uniqueness result for r as well.

8.1.8 Unique Factorization Domains

- We have shown that principal ideal domains have unique factorization. We now study the more general class of integral domains having unique factorization:
- Definition: An integral domain R is a unique factorization domain (UFD) if every nonzero nonunit $r \in R$ can be written as a finite product $r = p_1 p_2 \cdots p_d$ of irreducible elements, and this factorization is unique up to associates: if $r = p_1 p_2 \cdots p_d = q_1 q_2 \cdots q_k$ for irreducibles p_i and q_j, then $d = k$ and there is some reordering of the factors such that p_i is associate to q_i for each $1 \leq i \leq k$.
    - Example: As we proved in the previous section, every principal ideal domain is a unique factorization domain: thus $\mathbb{Z}$, $F[x]$, and $\mathbb{Z}[i]$ are unique factorization domains.
    - Example: As we essentially proved already (and will formally prove later) the polynomial ring $\mathbb{Z}[x]$ is a unique factorization domain, even though it is not a principal ideal domain.
    - There are two ways an integral domain can fail to be a unique factorization domain: one way is for some element to have two inequivalent factorizations, and the other way is for some element not to have any factorization.
    - Non-Example: The ring $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain because we can write $6 = (1 + \sqrt{-5})(1 - \sqrt{-5}) = 2 \cdot 3$. Note that each of $1 \pm \sqrt{-5}$, 2, and 3 is irreducible in $\mathbb{Z}[\sqrt{-5}]$ since their norms are 6, 4, and 9 respectively and there are no elements in $\mathbb{Z}[\sqrt{-5}]$ of norm 2 or 3, and none of these elements are associate to one another. Thus, 6 has two inequivalent factorizations into irreducibles in $\mathbb{Z}[\sqrt{-5}]$.
    - Non-Example: The ring $\mathbb{Z}[2i]$ is not a unique factorization domain because we can write $4 = 2 \cdot 2 = (2i) \cdot (2i)$. Note that both 2 and $2i$ are irreducible since their norms are both 4 and there are no elements in $\mathbb{Z}[2i]$ of norm 2, and 2 and $2i$ are not associate since $i \notin \mathbb{Z}[2i]$. Thus, 4 has two inequivalent factorizations into irreducibles in $\mathbb{Z}[2i]$.
    - Non-Example: The ring $\mathbb{Z} + x\mathbb{Q}[x]$ of polynomials with rational coefficients and integral constant term is not a unique factorization domain because not every element has a factorization. Explicitly, the element x is not irreducible since $x = 2 \cdot \frac{1}{2}x$ and neither 2 nor $\frac{1}{2}x$ is a unit, but x cannot be written as a finite product of irreducible elements: any such factorization would necessarily consist of a product of constants times a rational multiple of x, but no rational multiple of x is irreducible in $\mathbb{Z} + x\mathbb{Q}[x]$.
- Like in principal ideal domains, irreducible elements are the same as prime elements in unique factorization domains (and thus, we may interchangeably refer to "prime factorizations" or "irreducible factorizations" in a UFD):
- Proposition (Irreducibles are Prime in a UFD): Every irreducible element in a unique factorization domain is prime.
    - Proof: Suppose that p is an irreducible element of R and that $p|ab$ for some elements $a, b \in R$: we must show that $p|a$ or $p|b$.
    - Since R is a unique factorization domain, we may write $a = q_1 q_2 \cdots q_d$ and $b = r_1 r_2 \cdots r_k$ for some irreducibles q_i and r_j: then $q_1 q_2 \cdots q_d r_1 r_2 \cdots r_k = ab$. But since the factorization of ab into irreducibles is unique, p must be associate to one of the irreducibles q_i or r_j.
    - If p is associate to one of the q_i then $p|a$, while if p is associate to one of the r_j then $p|b$. Since at least one of these two must occur, $p|a$ or $p|b$, as required.
- Like in $\mathbb{Z}$, we can also describe greatest common divisors in terms of prime factorizations:
- Proposition (Divisibility in UFDs): If a and b are nonzero elements in a unique factorization domain R, then there exist units u and v and prime elements $p_1, p_2, \ldots, p_k$ no two of which are associate so that $a = up_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ and $b = vp_1^{b_1} p_2^{b_2} \cdots p_k^{b_k}$ for some nonnegative integers a_i and b_i. Furthermore, a divides b if and only if $a_i \leq b_i$ for all $1 \leq i \leq k$, and the element $d = p_1^{\min(a_1, b_1)} \cdots p_k^{\min(a_k, b_k)}$ is a greatest common divisor of a and b.
    - Proof: Since R is a UFD, we can write a as a product of irreducibles. As follows from a trivial induction, we can then "collapse" these factorizations by grouping together associates and factoring out the resulting units to obtain a factorization of the form $a = up_1^{a_1} p_2^{a_2} \cdots p_d^{a_d}$.

- We can repeat the process with b, and then add any further irreducibles that appear in its factorization to the end of the list, to obtain the desired factorizations $a = up_1^{a_1}p_2^{a_2}\cdots p_k^{a_k}$ and $b = vp_1^{b_1}p_2^{b_2}\cdots p_k^{b_k}$ for nonnegative integers a_i and b_i.
- For the statement about divisibility, if $a|b$ then we have $b = ar$ for some $r \in R$, so that $vp_1^{b_1}p_2^{b_2}\cdots p_k^{b_k} = up_1^{a_1}p_2^{a_2}\cdots p_k^{a_k}r$. But since p_i divides the right-hand side at least a_i times, we see that p_i must also divide the left-hand side at least a_i times: furthermore, since each of the terms excluding p_i is not associate to p_i, by a trivial induction we conclude that b_i must be at least as large as a_i, for each i.
- For the statement about the gcd, it is easy to see by the above that d divides both a and b. If d' is any other common divisor, then since d' divides a we see that any irreducible occurring in the prime factorization of d' must be associate to those appearing in the prime factorization of a, hence (by collapsing the factorization as above) we can write $d' = wp_1^{d_1}p_2^{d_2}\cdots p_k^{d_k}$ for some nonnegative integers d_i and some unit w.
- Then since d' is a common divisor of both a and b we see that $d_i \leq a_i$ and $d_i \leq b_i$, whence $d_i \leq \min(a_i, b_i)$ for each i: then d' divides d, so d is a greatest common divisor as claimed.

- We also recover one of the other fundamental properties of relatively prime elements and gcds:
- Corollary (Relatively Prime Elements and GCDs): In any unique factorization domain, d is a gcd of a and b if and only if a/d and b/d are relatively prime. Furthermore, if a and b are relatively prime and $a|bc$, then $a|c$.
  - Proof: Apply the previous proposition to write $a = up_1^{a_1}p_2^{a_2}\cdots p_k^{a_k}$ and $b = vp_1^{b_1}p_2^{b_2}\cdots p_k^{b_k}$ for some nonnegative integers a_i and b_i, irreducibles p_i, and units u and v.
  - Then $d = p_1^{\min(a_1,b_1)}\cdots p_k^{\min(a_k,b_k)}$ is a gcd of a and b, and it is easy to see that the exponent of p_i in a/d or b/d is zero for each i: thus, the only common divisors of a/d and b/d are units, so a/d and b/d are relatively prime.
  - Inversely, if $d' = wp_1^{d_1}p_2^{d_2}\cdots p_k^{d_k}$ is any other common divisor of a and b, and $d_i < \min(a_i, b_i)$ for some i, then p_i is a common divisor of a/d' and b/d' and thus the latter are not relatively prime.
  - For the second statement, consider the irreducible factors of bc: since a and b have no irreducible factors in common, every irreducible factor of c must divide a.

8.1.9 The Chinese Remainder Theorem

- As a last preliminary result, we give a general formulation of the Chinese remainder theorem. We first require a few preliminary definitions:
- Definition: If R is commutative with 1 and I and J are ideals of R, then the sum $I+J = \{a+b : a \in I, b \in J\}$ is defined to be the set of all sums of elements of I and J, and the product $IJ = \{a_1b_1 + \cdots + a_nb_n, : a_i \in I, b_i \in J\}$ is the set of finite sums of products of an element of I with an element of J.
  - It is not difficult to verify that $I + J$ and IJ are both ideals of R, and that IJ contains the intersection $I \cap J$.
  - If I and J are finitely generated, with $I = (a_1, a_2, \ldots, a_n)$ and $J = (b_1, b_2, \ldots, b_m)$, it is also not hard to see that $I + J = (a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m)$ and $IJ = (a_1b_1, a_1b_2, \ldots, a_1b_m, a_2b_1, \ldots, a_2b_m, \ldots, a_nb_m)$.
  - Example: If $I = (a)$ and $J = (b)$ inside $\mathbb{Z}$, then $I + J = (a, b) = (d)$ where $d = \gcd(a, b)$ and $IJ = (ab)$.
  - We can also speak of the product $I_1I_2\cdots I_n$ of more than two ideals, defined as the set of finite sums of products of an element from each of $I_1, I_2, \ldots, I_n$.
- Definition: If R is commutative with 1, the ideals I and J are comaximal if $I + J = R$.
  - Note that $a\mathbb{Z} + b\mathbb{Z} = \mathbb{Z}$ precisely when a and b are relatively prime. (The appropriate notion in general rings is not "primality" but "maximality", so we use the term comaximal rather than coprime.)
- We can now state the general Chinese remainder theorem:

- Theorem (Chinese Remainder Theorem): Let R be commutative with 1 and $I_1, I_2, \ldots, I_n$ be ideals of R. Then the map $\varphi : R \to (R/I_1) \times (R/I_2) \times \cdots \times (R/I_n)$ defined by $\varphi(r) = (r + I_1, r + I_2, \ldots, r + I_n)$ is a ring homomorphism with kernel $I_1 \cap I_2 \cap \cdots \cap I_n$. If all of the ideals $I_1, I_2, \ldots, I_n$ are pairwise comaximal, then φ is surjective and $I_1 \cap I_2 \cap \cdots \cap I_n = I_1 I_2 \cdots I_n$, and thus $R/(I_1 I_2 \cdots I_n) \cong (R/I_1) \times (R/I_2) \times \cdots \times (R/I_n)$.
    - Proof: First, φ is a homomorphism since $\varphi(a+b) = (a+b+I_1, \ldots, a+b+I_n) = (a+I_1, \ldots, a+I_n) + (b+I_1, \ldots, b+I_n) = \varphi(a) + \varphi(b)$ and similarly $\varphi(ab) = (ab+I_1, \ldots, ab+I_n) = (a+I_1, \ldots, a+I_n) \cdot (b+I_1, \ldots, b+I_n) = \varphi(a)\varphi(b)$.
    - The kernel of φ is the set of elements $r \in R$ such that $\varphi(r) = (0 + I_1, \ldots, 0 + I_n)$, which is equivalent to requiring $r \in I_1$, $r \in I_2$, ... , and $r \in I_n$: thus, $\ker \varphi = I_1 \cap I_2 \cap \cdots \cap I_n$.
    - For the second statement, we will prove the results for two ideals and then deduce the general statement via induction.
    - So suppose I and J are ideals of R and $\varphi : R \to (R/I) \times (R/J)$ has $\varphi(r) = (r + I, r + J)$. We must show that if $I + J = R$, then $I \cap J = IJ$ and φ is surjective.
    - If $I + J = R$ then by definition there exist elements $x \in I$ and $y \in J$ with $x + y = 1$.
    - Then for any $r \in I \cap J$, we can write $r = r(x + y) = rx + yr$, and both rx and yr are in IJ: hence $I \cap J \subseteq IJ$, and since $IJ \subseteq I \cap J$ we conclude $IJ = I \cap J$.
    - Furthermore, for any $a, b \in R$ we can write $ay + bx = a(1 - x) + bx = a + (b - a)x$ so $ay + bx \in a + I$, and likewise $ay + bx = ay + b(1 - y) = b + (a - b)y \in b + J$.
    - Then $\varphi(ay + bx) = (ay + bx + I,\ ay + bx + J) = (a + I,\ b + J)$, and therefore φ is surjective as claimed.
    - Finally, the statement that $R/IJ \cong (R/I) \times (R/J)$ then follows immediately by the first isomorphism theorem. This establishes all of the results for two ideals.
    - For the general statement, we use induction on n: the base case $n = 2$ was done above, and for the inductive step, it is enough to show that the ideals I_1 and $I_2 \cdots I_n$ are comaximal, since then we may write $R/(I_1 I_2 \cdots I_n) \cong (R/I_1) \times (R/I_2 \cdots I_n)$ and apply the induction hypothesis to $R/I_2 \cdots I_n$.
    - If I_1 and I_i are comaximal for $2 \leq i \leq n$, then there exist elements $x_i \in I_1$ and $y_i \in I_i$ such that $x_i + y_i = 1$. Then $1 = (x_2 + y_2)(x_3 + y_3) \cdots (x_n + y_n) \equiv y_2 y_3 \cdots y_n$ modulo I_1. But since $y_2 y_3 \cdots y_n$ is in $I_2 I_3 \cdots I_n$, this means that $I_1 + I_2 I_3 \cdots I_n$ contains 1 and is therefore all of R, as required.
- The name of this theorem comes from its application inside $\mathbb{Z}$ to solving simultaneous modular congruences.
    - Explicitly, if $m_1, m_2, \ldots m_n$ are relatively prime positive integers, then $\varphi : \mathbb{Z} \to (\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/m_n\mathbb{Z})$ given by $\varphi(a) = (a \bmod m_1,\ a \bmod m_2, \ldots, a \bmod m_n)$ is a surjective homomorphism with kernel $m_1 m_2 \cdots m_n \mathbb{Z}$.
    - The fact that this map is surjective says that the system of simultaneous congruences $x \equiv a_1 \bmod m_1$, $x \equiv a_2 \bmod m_2$, ... , $x \equiv a_n \bmod m_n$ always has a solution in $\mathbb{Z}$. Furthermore, the characterization of the kernel says that the solution is unique modulo $m_1 m_2 \cdots m_n$.
    - Systems of congruences of this form were studied by the ancient Chinese, whence the theorem's name.
- A useful application of the Chinese remainder theorem is to decompose $\mathbb{Z}/m\mathbb{Z}$ as the direct product of other rings when m is composite. This particular application is the generalization of the classical version of the Chinese remainder theorem as applied to integer congruences:
- Corollary (Chinese Remainder Theorem for $\mathbb{Z}$): If m is a positive integer with prime factorization $m = p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n}$, then $\mathbb{Z}/m\mathbb{Z} \cong (\mathbb{Z}/p_1^{a_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_n^{a_n}\mathbb{Z})$.
    - Remark: By counting the number of units in the Cartesian product, we see that the number of units in $\mathbb{Z}/m\mathbb{Z}$ is $m(1 - 1/p_1)(1 - 1/p_2) \cdots (1 - 1/p_n)$: this gives us an alternate derivation of the formula for the Euler φ-function $\varphi(m)$.
    - Proof: This statement follows from the Chinese remainder theorem along with the observation that if p and q are distinct primes, then the ideals $p^a\mathbb{Z}$ and $q^b\mathbb{Z}$ are comaximal in $\mathbb{Z}$.

8.2 Factorization In Quadratic Integer Rings[2]

- We now turn our attention to the question of factorization in quadratic integer rings.

8.2.1 Unique Factorization of Elements in $\mathcal{O}_D$

- As we have seen, some of the quadratic integer rings (like $\mathbb{Z}[i]$) are unique factorization domains, while others (like $\mathbb{Z}[\sqrt{-5}]$) are not.
  - More specifically, by extending the argument used for $\mathbb{Z}[i]$, it can be shown that the quadratic integer ring $\mathcal{O}_D = \mathcal{O}_{\mathbb{Q}(\sqrt{D})} = \begin{cases} \mathbb{Z}[\sqrt{D}] & \text{for } D \equiv 2,3 \pmod 4 \\ \mathbb{Z}[(1+\sqrt{D})/2] & \text{for } D \equiv 1 \pmod 4 \end{cases}$ is Euclidean (with norm given by the field norm) for a known list of negative $D = -1, -2, -3, -7, -11$ and for various positive D, including $D = 2, 3, 5, 6, 7, 11, \ldots$.
  - We would like to know whether it is possible to recover some sort of "unique factorization" property in the quadratic integer rings, even when they are not unique factorization domains.
- The question of when $\mathcal{O}_D$ is a UFD was (and is) of substantial interest in applications to solving equations in number theory, since we may use properties of integer rings (e.g., $\mathbb{Z}[i]$) to characterize the solutions to such equations, as we saw earlier in the case of the equation $a^2 + b^2 = c^2$.
  - For example, if p is an odd prime, we may study the Fermat equation $x^p + y^p = z^p$ in the ring $\mathbb{Z}[\zeta_p] = \{a_0 + a_1\zeta_p + \cdots + a_{p-1}\zeta_p^{p-1} : a_i \in \mathbb{Z}\}$ where $\zeta_p = e^{2\pi i/p} = \cos(2\pi/p) + i\sin(2\pi/p)$ is a nonreal pth root of unity (satisfying $\zeta_p^p = 1$).
  - We may rearrange the equation as $z^p - y^p = x^p$ and then factor the left-hand side as the product $(z-y)(z-\zeta_p y)(z-\zeta_p^2 y)\cdots(z-\zeta_p^{p-1}y)$ of linear terms inside $\mathbb{Z}[\zeta_p]$.
  - If $\mathbb{Z}[\zeta_p]$ were a unique factorization domain, then since the terms on the left-hand side are essentially relatively prime, each of them would have to be a pth power in $\mathbb{Z}[\zeta_p]$, up to some small factors. But this can be shown not to be possible unless $y = 0$, and so we would be able to conclude that Fermat's equation $x^p + y^p = z^p$ has no nontrivial integer solutions.
  - Unfortunately, the ring $\mathbb{Z}[\zeta_p]$ is not always a unique factorization domain. But the study of Diophantine equations in number theory, and associated questions about unique factorization, were (historically speaking) the impetus for much of the development of modern algebra, including ring theory.
- We will restrict our attention to quadratic integer rings, since we can give concrete arguments in these cases. For example, we can show that every element does possess at least one factorization (and thus, the failure to be a UFD lies entirely with non-uniqueness):
- Proposition (Element Factorizations in $\mathcal{O}_D$): If $R = \mathcal{O}_D$ is a quadratic integer ring, then every nonzero nonunit in R has at least one factorization as a product of irreducible elements.
  - Proof: We show the result by (strong) induction on the absolute value of the norm $N(r)$. If $N(r) = 0$ then $r = 0$, while if $N(r) = \pm 1$ then r is a unit.
  - For the base case we take $|N(r)| = 2$: then r is irreducible, since the absolute value of its norm is a prime. (This follows by the same argument used in $\mathbb{Z}[i]$.)
  - For the inductive step, suppose that $|N(r)| = n$ for $n \geq 3$. If r is irreducible we are done: otherwise we have $r = ab$ for some a, b with $1 < |N(a)|, |N(b)| < n$.
  - By the inductive hypothesis, both a and b have factorizations as a product of irreducibles, so r does too.
- It would appear that we are essentially at an impasse regarding factorization of elements. However, if we shift our focus instead to ideals, it turns out that these rings do possess unique prime factorizations on the level of *ideals*, rather than elements.

[2] The treatment of some of the material in this section is adapted from notes of Keith Conrad: `http://www.math.uconn.edu/~kconrad/blurbs/gradnumthy/quadraticgrad.pdf`.

- In fact, this is where the name "ideal" originally arose: in Kummer's study of unique factorization, he constructed "ideal numbers" (essentially as sets of linear combinations of elements of $\mathcal{O}_D$) and proved that they did possess unique prime factorization. These "ideal numbers" were the prototype of the modern definition of an ideal.
- To illustrate using an example we have already discussed, the element $6 \in \mathbb{Z}[\sqrt{-5}]$ has two different factorizations into irreducibles, as $2 \cdot 3 = 6 = (1+\sqrt{-5}) \cdot (1-\sqrt{-5})$.
- This yields the equivalent ideal factorization $(6) = (2) \cdot (3) = (1+\sqrt{-5}) \cdot (1-\sqrt{-5})$.
- However, as ideals, we can factor further: explicitly, one can verify that $(2) = (2, 1+\sqrt{-5})^2$, that $(1 \pm \sqrt{-5}) = (2, 1+\sqrt{-5}) \cdot (3, 1 \pm \sqrt{-5})$, and that $(3) = (3, 1+\sqrt{-5}) \cdot (3, 1-\sqrt{-5})$.
- For an example of one of these calculations: we have $(2, 1+\sqrt{-5}) \cdot (3, 1+\sqrt{-5}) = (6, 2+2\sqrt{-5}, 3+3\sqrt{-5}, -4+2\sqrt{-5})$. We can reduce the generating set by observing that this ideal contains $(3+3\sqrt{-5}) - (2+2\sqrt{-5}) = 1+\sqrt{-5}$, and that each of the four generators of the product ideal is a multiple of $1+\sqrt{-5}$: thus, in fact, $(2, 1+\sqrt{-5}) \cdot (3, 1+\sqrt{-5}) = (1+\sqrt{-5})$, as claimed. The other calculations are similar.
- On the level of ideals, therefore, we see that these two factorizations are really "the same": both of them reduce to the factorization $(6) = (2, 1+\sqrt{-5})^2 \cdot (3, 1+\sqrt{-5}) \cdot (3, 1-\sqrt{-5})$.
- Furthermore, each of the ideals $(2, 1+\sqrt{-5})$, $(3, 1+\sqrt{-5})$, and $(3, 1-\sqrt{-5})$ can be shown to be prime (the quotient ring of $\mathbb{Z}[\sqrt{-5}]$ by each is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/3\mathbb{Z}$, and $\mathbb{Z}/3\mathbb{Z}$ respectively).
- Thus, we have found a factorization of the ideal (6) as a product of prime ideals of $\mathbb{Z}[\sqrt{-5}]$.

- Our goal is to show that the behavior in the example above holds in general: namely, that we can write any nonzero ideal in a quadratic integer ring as a product of prime ideals, and that this factorization is unique up to rearrangement.
  - After first establishing some important properties of prime ideals, our model will be similar to our proofs that $\mathbb{Z}$ and $F[x]$ have unique factorization: we will discuss some properties of divisibility, show that every nonzero ideal can be written as a product of prime ideals, and then show that the factorization is unique.
  - We will then give some applications of unique factorization into prime ideals, and in particular describe how to compute the prime ideals of $\mathcal{O}_D$.

8.2.2 Ideals in $\mathcal{O}_D$

- To begin, we show that every ideal in $\mathcal{O}_D$ is generated by at most 2 elements:
- Proposition (Ideal Generators in $\mathcal{O}_D$): If $R = \mathcal{O}_D$ is a quadratic integer ring, then every ideal in R is of the form $(n, a + b \cdot \frac{1+\sqrt{D}}{2})$ for some $a, b, n \in \mathbb{Z}$.
  - Proof: Let I be an ideal of $\mathcal{O}_D$, and define $I_0 = I \cap \mathbb{Z}$ and I_1 to be the set of $r \in \mathbb{Z}$ such that there exists some $s \in \mathbb{Z}$ with $s + r \cdot \frac{1+\sqrt{D}}{2} \in I$.
  - Observe that I_0 and I_1 are both ideals of $\mathbb{Z}$ since they clearly contain 0, are closed under subtraction, and are closed under arbitrary $\mathbb{Z}$-multiplication. So suppose $I_0 = (n)$ and $I_1 = (b)$: then $n \in I$, and by definition of I_1, there exists $a \in \mathbb{Z}$ such that $a + b \cdot \frac{1+\sqrt{D}}{2} \in I$.
  - We claim that n and $a + b \cdot \frac{1+\sqrt{D}}{2}$ generate I, so suppose $s + r \cdot \frac{1+\sqrt{D}}{2}$ is an arbitrary element of I. By definition of I_1 we see that $r \in I_1$, whence $r = yb$ for some $y \in \mathbb{Z}$.
  - Then $\left[\left(s + r \cdot \frac{1+\sqrt{D}}{2}\right) - y \cdot \left(a + b \cdot \frac{1+\sqrt{D}}{2}\right)\right] = s - ay$ is in $I \cap \mathbb{Z} = I_0$, so this quantity is equal to xn for some $x \in \mathbb{Z}$.
  - Thus, $s + r \cdot \frac{1+\sqrt{D}}{2} = xn + y\left(a + b \cdot \frac{1+\sqrt{D}}{2}\right)$, and so n and $a + b \cdot \frac{1+\sqrt{D}}{2}$ generate I as claimed.

- As a corollary, we can deduce that nonzero prime ideals of $\mathcal{O}_D$ are maximal:
- Corollary (Quotients of $\mathcal{O}_D$): If $R = \mathcal{O}_D$ is a quadratic integer ring and I is a nonzero ideal, then $\mathcal{O}_D/I$ is finite. In particular, every nonzero prime ideal of $\mathcal{O}_D$ is maximal.
  - Proof: For the first statement, if I is a nonzero ideal in $\mathcal{O}_D$, then $I \cap \mathbb{Z}$ is nonzero (since if $r \in I$ is any nonzero element, $N(r) \in I$ is a nonzero integer) and so by the proposition above, $I = (n, a + b \cdot \frac{1+\sqrt{D}}{2})$ where $n \neq 0$ is a generator of $I \cap \mathbb{Z}$.
  - There are finitely many residue classes in $\mathcal{O}_D/(n)$, since each residue class has (exactly) one representative by an element of the form $s + t \cdot \frac{1+\sqrt{D}}{2}$ for some integers $0 \leq s, t \leq n-1$. Then by the third isomorphism theorem, we know that $\mathcal{O}_D/I \cong [\mathcal{O}_D/(n)]/[I/(n)]$ is a quotient of a finite ring, hence also finite.
  - For the second statement, if P is a nonzero prime ideal of $\mathcal{O}_D$, then $\mathcal{O}_D/P$ is a finite integral domain, hence is a field.
- We also require a few additional properties about the "conjugation" map in $\mathcal{O}_D$:
- Definition: If $a + b\sqrt{D}$ is an element of $\mathcal{O}_D$, its conjugate is $\overline{a + b\sqrt{D}} = a - b\sqrt{D}$. For any $r \in \mathcal{O}_D$, we have $N(r) = r \cdot \overline{r}$, and we also define the trace of r as $\text{tr}(r) = r + \overline{r}$.
  - It is not hard to see that both $N(r)$ and $\text{tr}(r)$ are elements of $\mathbb{Z}$ for any $r \in \mathcal{O}_D$.
  - Conversely, the elements $r \in \mathbb{Q}(\sqrt{D})$ with the property that $N(r)$ and $\text{tr}(r)$ are both in $\mathbb{Z}$ are precisely the elements of $\mathcal{O}_D$.
  - To see this, if $r = a + b\sqrt{D} \in \mathbb{Q}(\sqrt{D})$, then $N(r) = a^2 - Db^2$ and $\text{tr}(r) = 2a$. If both of these values are integers, then $2a$ is an integer, and then $4N(r) - (2a)^2 = -4Db^2$ is also an integer. Since D is squarefree, this means $4b^2$ hence $2b$ is an integer as well.
  - Finally, if $D \equiv 2, 3 \pmod 4$ then $N(r)$ will only be an integer when a and b are themselves integers, while if $D \equiv 1 \pmod 4$ then $N(r)$ will be an integer when $2a$ and $2b$ are integers of the same parity. In both cases, we see $r \in \mathcal{O}_D$ as claimed.
- Definition: If I is an ideal of $\mathcal{O}_D$, then its conjugate is the ideal $\overline{I} = \{\overline{r} : r \in I\}$.
  - It is easy to see that if $I = (r, s)$, then $\overline{I} = (\overline{r}, \overline{s})$, so for example in $\mathbb{Z}[\sqrt{-5}]$ we have $\overline{(3, 1 + \sqrt{-5})} = (3, 1 - \sqrt{-5})$.
  - Likewise, it is a straightforward calculation that for any ideals I and J, we have $\overline{IJ} = \overline{I} \cdot \overline{J}$ and $\overline{\overline{I}} = I$.
- Our first key result is that the product of an ideal with its conjugate is always principal:
- Theorem (Ideals and Conjugates in $\mathcal{O}_D$): If I is any ideal of $\mathcal{O}_D$, then $I \cdot \overline{I}$ is always principal.
  - Proof: If $I = 0$ we are done. Otherwise, suppose that $I = (r, s)$ for some nonzero $r, s \in \mathcal{O}_D$: then $\overline{I} = (\overline{r}, \overline{s})$ and $I \cdot \overline{I} = (r\overline{r}, r\overline{s}, \overline{r}s, s\overline{s})$.
  - We claim in fact that $I \cdot \overline{I} = (r\overline{r}, r\overline{s} + \overline{r}s, s\overline{s}) = (N(r), \text{tr}(r\overline{s}), N(s))$.
  - To see this, observe that $N(r)$, $\text{tr}(r\overline{s})$, and $N(s)$ are each in $\mathbb{Z}$, so let their greatest common divisor be d. Then $d = xN(r) + y\text{tr}(r\overline{s}) + zN(s)$ for some $x, y, z \in \mathbb{Z}$, and so $(N(r), \text{tr}(r\overline{s}), N(s)) = (d)$ in $\mathcal{O}_D$.
  - In order to show that $I \cdot \overline{I} = (r\overline{r}, r\overline{s} + \overline{r}s, s\overline{s})$, we must show that $r\overline{s}$ is in the ideal $(r\overline{r}, r\overline{s} + \overline{r}s, s\overline{s}) = (d)$.
  - Observe that $\text{tr}(r\overline{s}/d) = \frac{r\overline{s} + \overline{r}s}{d} = \frac{\text{tr}(r\overline{s})}{d}$ is an integer, as is $N(r\overline{s}/d) = \frac{r\overline{s}}{d} \cdot \frac{\overline{r}s}{d} = \frac{N(r)}{d} \cdot \frac{N(s)}{d}$, since d divides each of $N(r)$, $\text{tr}(r\overline{s})$, and $N(s)$.
  - Then, by our characterization of the elements in $\mathcal{O}_D$, we conclude that $r\overline{s}/d$ is in $\mathcal{O}_D$, so that $r\overline{s} \in (d)$.
  - Therefore, $I \cdot \overline{I} = (r\overline{r}, r\overline{s} + \overline{r}s, s\overline{s}) = (N(r), \text{tr}(r\overline{s}), N(s)) = (d)$ is principal, as claimed.

8.2.3 Divisibility and Unique Factorization of Ideals in $\mathcal{O}_D$

- Next, we discuss divisibility of ideals.
- Definition: If I and J are ideals of $\mathcal{O}_D$, we say that I divides J, written $I|J$, if there is some ideal K such that $J = IK$.
- Proposition (Properties of Ideal Divisibility): Suppose I and J are ideals of $\mathcal{O}_D$ and $r \in \mathcal{O}_\mathcal{D}$.
  1. If I divides J, then I contains J.
     - Proof: If $J = IK$ then every element in J is a sum of multiples of elements in I, hence is in I.
  2. We have $I|J$ and $J|I$ if and only if $I = J$.
     - Proof: Since $I = IR$, $I = J$ implies $I|J$ and $J|I$. Conversely, if $I|J$ and $J|I$, then $I \subseteq J$ and $J \subseteq I$ so $I = J$.
  3. The principal ideal (r) divides I if and only if (r) contains I.
     - Proof: The forward direction follows from (1). For the reverse, if (r) contains $I = (s, t)$ then $r|s$ and $r|t$, and then $I = (r) \cdot (s/r, t/r)$.
  4. If $(r)J = (r)K$ and $r \neq 0$, then $J = K$.
     - Proof: If $s \in J$, then $rs \in (r)J$: then $rs \in (r)K$ and so $s \in K$. Thus, $J \subseteq K$, and by the same argument in reverse, $K \subseteq J$, so $J = K$.
  5. If $IJ = IK$ and $I \neq 0$, then $J = K$.
     - Proof: If $I \neq 0$ then $I \cdot \overline{I} = (r)$ is a nonzero principal ideal as we proved above. Then $IJ = IK$ implies $(I\overline{I})J = (I\overline{I})K$ so that $(r)J = (r)K$, whence $J = K$ by (4).
  6. The ideal I divides J if and only if I contains J.
     - Proof: The forward direction is given by (1), and it is easy to see that the result also holds if I is zero (since every ideal divides the zero ideal, but the zero ideal only divides itself).
     - If I and J are nonzero ideals and I contains J, then $I \cdot \overline{I} = (r)$ contains $J \cdot \overline{I}$.
     - Then by (3) we see that $(r) = I \cdot \overline{I}$ divides $J \cdot \overline{I}$, so $J \cdot \overline{I} = I \cdot \overline{I} \cdot K$ for some K. Then since $I \neq 0$ (whence $\overline{I} \neq 0$), by (5) we may cancel to conclude that $J = IK$, meaning that I divides J.
- The upshot of the previous proposition is that dividing is the same as containment, on the level of ideals.
  - From this description and the fact that nonzero prime ideals are maximal, we can immediately conclude that the "irreducible" ideals (namely, ideals that have no nontrivial factorization, which is to say $I = JK$ implies $J = \mathcal{O}_D$ or $K = \mathcal{O}_D$) are the same as the maximal ideals, which are in turn the same as the nonzero prime ideals.
- It remains for us to establish that every nonzero ideal has a factorization into prime ideals, and that the factorization is unique.
  - To show that nonzero ideals have a factorization, we will mimic the proof we gave earlier for elements by defining an "ideal norm".
  - For elements we use the norm $N(r) = |r \cdot \overline{r}|$, so a natural guess for ideals would be to use $I \cdot \overline{I}$: conveniently enough, we have proven that this ideal is principal and generated by an integer.
- Definition: If I is an ideal of $\mathcal{O}_D$, then the norm $N(I)$ of I is the nonnegative integer generator of the principal ideal $I \cdot \overline{I}$.
  - Observe that the norm is multiplicative: $(N(IJ)) = IJ \cdot \overline{IJ} = I\overline{I} \cdot J\overline{J} = (N(I)N(J))$.
  - Also notice that the only ideal with norm 0 is the zero ideal, while the only ideal with norm 1 is $\mathcal{O}_D$ (since $I\overline{I} = (1)$ implies that I contains a unit).
  - Thus, in particular, if $N(I)$ is a prime integer then I has no nontrivial factorization, and thus I is a prime ideal.

- We can now establish that every ideal has a factorization as a product of prime ideals:
- Proposition (Prime Factorization of Ideals in $\mathcal{O}_D$): Every nonzero ideal in $\mathcal{O}_D$ can be written as the product of prime ideals of $\mathcal{O}_D$.
    - As usual, we take the convention that the empty product represents $\mathcal{O}_D$.
    - Proof: We use (strong) induction on the norm of the ideal. Since $I \neq 0$ we have $N(I) \geq 1$.
    - For the base case $N(I) = 1$, we have $I = \mathcal{O}_D$ so we may take the empty product of prime ideals.
    - For the inductive step, suppose the result holds for every ideal of norm less than n and suppose $N(I) = n$.
    - If I is a prime ideal we are done, so assume I is not prime (hence not maximal). Then I is properly contained in some other proper ideal J, so by our results on divisibility we may write $I = JK$ where J and K are both proper.
    - Then $N(I) = N(J) \cdot N(K)$ and $1 < N(J), N(K) < n$. By the inductive hypothesis, both J and K are the product of some number of prime ideals, so I is as well.
- As our final step, we show that the factorization is unique. To do this we require the prime divisibility property of prime ideals:
- Proposition (Divisibility and Prime Ideals in $\mathcal{O}_D$): If P is a prime ideal of $\mathcal{O}_D$ and I and J are any ideals with $P|IJ$, then $P|I$ or $P|J$.
    - Proof: By the equivalence of divisibility and containment in $\mathcal{O}_D$, we need to show that if P is a prime ideal with P containing IJ, then P contains I or P contains J.
    - Suppose that P contains neither I nor J: then there is some $x \in I$ that is not in P and some $y \in J$ that is not in P. But then $xy \in IJ$ is contained in P, contradicting the assumption that P was prime. Thus, P contains I or P contains J, as required.
- Theorem (Uniqueness of Prime Ideal Factorization in $\mathcal{O}_D$): Every nonzero ideal in $\mathcal{O}_D$ can be written as the product of prime ideals of $\mathcal{O}_D$. Furthermore, this representation is unique up to rearrangement: if $I = P_1P_2 \cdots P_n = Q_1Q_2 \cdots Q_k$, then $n = k$ and there is some rearrangement of the Q_i so that $P_i = Q_i$.
    - Proof: We proved above that every nonzero ideal can be written as a product of prime ideals.
    - For the uniqueness, we induct on the minimal number of terms n in the prime factorization.
    - For the base case $n = 0$, we have $I = \mathcal{O}_D$: since every prime ideal is proper, we cannot write I as a nonempty product of prime ideals.
    - For the inductive step, suppose that every representation with fewer than n terms is unique, and suppose $I = P_1P_2 \cdots P_n = Q_1Q_2 \cdots Q_k$. Since P_1 is prime and divides $Q_1Q_2 \cdots Q_k$, we see that P_1 must divide one of the Q_i; without loss of generality, rearrange so that P_1 divides Q_1.
    - But since P_1 and Q_1 are both nonzero prime ideals, they are maximal. Since P_1 divides Q_1 we see that P_1 contains Q_1, hence since Q_1 is maximal and $P_1 \neq \mathcal{O}_D$, we must have $P_1 = Q_1$.
    - Then by our ideal divisibility properties, we may cancel to obtain $P_2 \cdots P_n = Q_2 \cdots Q_k$, which by the inductive hypothesis has a unique factorization. Thus, the factorization of I is unique as claimed.

8.2.4 Calculating Factorizations in $\mathcal{O}_D$

- As a corollary of the unique factorization of ideals, we can give a characterization of when $\mathcal{O}_D$ is a unique factorization domain:
- Theorem (Unique Factorization in $\mathcal{O}_D$): The ring $\mathcal{O}_D$ is a unique factorization domain if and only if it is a principal ideal domain.
    - Inversely, this says that every example of non-unique factorization of elements in $\mathcal{O}_D$ ultimately arises from the presence of nonprincipal ideals.
    - Proof: Every PID is a UFD, so we need only prove the forward direction, so suppose $\mathcal{O}_D$ is a unique factorization domain.

- First suppose that P is a prime ideal: then P divides the principal ideal $(N(P))$. By the unique factorization of elements in $\mathcal{O}_D$, we can write $N(P) = \pi_1\pi_2\cdots\pi_n$ for some irreducible elements $\pi_1, \ldots, \pi_n \in \mathcal{O}_D$.
- Therefore, P divides the ideal product $(N(P)) = (\pi_1)\cdots(\pi_n)$, and hence P divides one of the ideals (π_i).
- But since irreducibles are prime in UFDs, the ideal (π_i) is also prime, and so we must have $P = (\pi_i)$, and so in particular P is principal.
- Then any nonzero ideal in $\mathcal{O}_D$ is a product of prime (hence principal) ideals hence is also principal. Since the zero ideal is also principal, every ideal in $\mathcal{O}_D$ is principal, so it is a PID.

- We can also describe how prime ideals in $\mathcal{O}_D$ arise in a more concrete way:

- Proposition (Prime Ideals in $\mathcal{O}_D$): If P is a nonzero prime ideal of $\mathcal{O}_D$, then $P \cap \mathbb{Z} = p\mathbb{Z}$ for a unique prime $p \in \mathbb{Z}$ (we say P "lies above" the prime ideal $p\mathbb{Z}$ of $\mathbb{Z}$). Furthermore, every prime ideal in $\mathcal{O}_D$ lying above $p\mathbb{Z}$ divides the ideal (p) in $\mathcal{O}_D$, and the norm of any prime ideal is either p or p^2.
  - Proof: Let $\varphi : \mathbb{Z} \to \mathcal{O}_D$ be the inclusion homomorphism, and observe that $\varphi^{-1}(P) = P \cap \mathbb{Z}$ is then an ideal of $\mathbb{Z}$, since the inverse image contains 0 and is closed under subtraction and arbitrary multiplication.
  - Furthermore, if $ab \in \varphi^{-1}(P)$ then $\varphi(a)\varphi(b) = \varphi(ab) \in P$, so since P is prime we see $\varphi(a) \in P$ or $\varphi(b) \in P$: thus, either a or b is in $\varphi^{-1}(P)$. Furthermore, since φ maps $1_{\mathbb{Z}}$ to $1_{\mathcal{O}_D}$, $\varphi^{-1}(P)$ does not contain 1, and since P contains the nonzero integer $N(P)$, we conclude that $\varphi^{-1}(P) = P \cap \mathbb{Z}$ is a nonzero prime ideal of $\mathbb{Z}$.
  - Then $P \cap \mathbb{Z} = p\mathbb{Z}$ for a unique prime $p \in \mathbb{Z}$. Thus, P contains $p \in \mathbb{Z}$ hence P contains (p), so by the equivalence of divisibility and containment, we see that P divides (p).
  - For the last statement, since P divides (p) we see that $N(P)$ divides $N((p)) = N(p) = p^2$, so since $N(P) > 1$ we must have $N(p) = p$ or $N(p) = p^2$.

- The result above tells us that we can find all the prime ideals in $\mathcal{O}_D$ by studying the factorization of the ideal (p) in $\mathcal{O}_D$.
  - Indeed, we have already seen how this works when $\mathcal{O}_D = \mathbb{Z}[i]$: there is a unique prime ideal $(1+i)$ above 2, with $(2) = (1+i)^2$ decomposing as a product with repeated factors, if $p \equiv 3 \bmod 4$ then the ideal (p) remains prime in $\mathbb{Z}[i]$, and if $p \equiv 1 \bmod 4$ then $(p) = (\pi)(\overline{\pi})$ factors as the product of distinct ideals.
  - We can recast this characterization as follows: if the polynomial $x^2 + 1$ has a repeated root modulo p (which only happens with $p = 2$) then the ideal (p) decomposes as a product with repeated factors, if $x^2 + 1$ remains irreducible modulo p (which is equivalent to saying that -1 is not a square modulo p, which occurs when $p \equiv 3 \bmod 4$) then (p) remains prime in $\mathbb{Z}[i]$, and if $x^2 + 1$ factors with distinct terms modulo p (which is equivalent to saying that -1 is a square modulo p, which occurs when $p \equiv 1 \bmod 4$) then (p) factors as the product of two distinct conjugate ideals.

- We can establish a similar characterization for the prime ideals of $\mathcal{O}_D$, which is a special case of a general result known as the Dedekind-Kummer factorization theorem:

- Theorem (Factorization of (p) in $\mathcal{O}_D$): Let p be a prime and let $q(x) = \begin{cases} x^2 - D & \text{for } D \equiv 2, 3 \bmod 4 \\ x^2 - x - (D-1)/4 & \text{for } D \equiv 1 \bmod 4 \end{cases}$, where $\omega = \begin{cases} \sqrt{D} & \text{for } D \equiv 2, 3 \bmod 4 \\ (1+\sqrt{D})/2 & \text{for } D \equiv 1 \bmod 4 \end{cases}$ is a root of $q(x)$. If the polynomial $q(x)$ has a repeated root r modulo p then the ideal $(p) = (p, \omega - r)^2$ is the square of a prime ideal of norm p in $\mathcal{O}_D$, if $q(x)$ is irreducible modulo p then the ideal (p) is prime in $\mathcal{O}_D$ of norm p^2, and if $q(x)$ is reducible with distinct roots r, r' modulo p, then $(p) = (p, \omega - r) \cdot (p, \omega - r')$ factors as the product of two distinct ideals in $\mathcal{O}_D$ each of norm p.
  - We note that $q(x)$ has a root modulo p if and only if D is a square modulo p. Also, $q(x)$ has a repeated root when $p|D$ (for any D) or when $p = 2$ and $D \equiv 3 \bmod 4$.
  - Proof: First observe that $\mathcal{O}_D \cong \mathbb{Z}[x]/(q(x))$, so by the isomorphism theorems we see that $\mathcal{O}_D/(p) \cong [\mathbb{Z}[x]/(q(x))]/(p) \cong \mathbb{Z}[x]/(p, q(x)) \cong [\mathbb{Z}[x]/(p)]/(q(x)) \cong \mathbb{F}_p[x]/(q(x))$. Thus, the ring structure of $\mathcal{O}_D/(p)$ is the same as the ring structure of $\mathbb{F}_p[x]/(q(x))$.

- The ideal (p) is prime (equivalently, maximal) in $\mathcal{O}_D$ precisely when the quotient ring is a field, and this occurs exactly when $q(x)$ is irreducible in $\mathbb{F}_p[x]$. In this case, $N((p)) = p^2$ so (p) is prime of norm p^2.
- If (p) is not prime, then since $N((p)) = p^2$, we see that (p) must factor as the product of two prime ideals I and I' each of norm p. Furthermore, since $I \cdot \overline{I} = (N(I)) = (p)$, by uniqueness of the prime ideal factorization we see that $I' = \overline{I}$, so the ideals in the factorization are conjugates.
- If $I \neq \overline{I}$ then $I + \overline{I} = \mathcal{O}_D$, so I and $\overline{I}$ are comaximal: then by the Chinese remainder theorem see that $\mathcal{O}_D/(p) \cong \mathcal{O}_D/I \times \mathcal{O}_D/\overline{I}$ is the direct product of two fields, and has no nonzero nilpotent elements.
- On the other hand, if $I = \overline{I}$, then $\mathcal{O}_D/(p) = \mathcal{O}_D/I^2$ has a nonzero nilpotent element (namely, the class of any element in I but not in I^2).
- For the other side, if $q(x) = (x-r)(x-r')$ in $\mathbb{F}_p[x]$, then the quotient ring $\mathcal{O}_D/(p) \cong \mathbb{F}_p[x]/(q(x)) \cong \mathbb{F}_p[x]/(x-r) \times \mathbb{F}_p[x]/(x-r') \cong \mathbb{F}_p \times \mathbb{F}_p$ is a direct product of two fields by the Chinese remainder theorem, and has no nonzero nilpotent elements.
- If $q(x) = (x-r)^2$ in $\mathbb{F}_p[x]$, then $\mathcal{O}_D/(p) \cong \mathbb{F}_p[x]/(x-r)^2$ does have a nonzero nilpotent element (namely $x - r$).
- Thus, comparing the ring structures in the two cases immediately shows that the case where $I = \overline{I}$ corresponds to the case where $q(x)$ has a repeated root, and $I \neq \overline{I}$ corresponds to the case where $q(x)$ has distinct roots.
- For the remaining statements, if r is a root of $q(x)$ in $\mathbb{F}_p$, then $(p, \omega - r)$ divides (p) since it contains (p), and since $\omega - r \notin (p)$ we see that $(p, \omega - r)$ is a proper divisor of (p).
- Furthermore, $N((p, \omega - r))$ is the greatest common divisor of $N(p) = p^2$, $\mathrm{tr}(p(\omega - r)) = p\mathrm{tr}(\omega - r)$, and $N(\omega - r) = q(r) \equiv 0 \bmod p$. Since each of the terms is divisible by p, the gcd cannot be 1, and therefore $(p, \omega - r)$ is a proper ideal. By the uniqueness of the prime ideal factorization, we conclude that $(p, \omega - r)$ must be a prime ideal dividing (p).
- If (p) is the square of a prime ideal, we then see $(p) = (p, \omega - r)^2$, while if (p) is the product of distinct ideals, we see that (p) is divisible by both $(p, \omega - r)$ and $(p, \omega - r')$, and since these ideals are comaximal we conclude $(p) = (p, \omega - r) \cdot (p, \omega - r')$. This establishes everything, so we are done.

- Example: Find the prime ideal factorizations of (2), (3), (5), and (7) in $\mathcal{O}_7 = \mathbb{Z}[\sqrt{7}]$.
  - For (2) we consider $x^2 - 7$ modulo 2: since it has a repeated root 1, we see $(2) = (2, \sqrt{7} - 1)^2$ in $\mathbb{Z}[\sqrt{7}]$.
  - For (3) we consider $x^2 - 7$ modulo 3: since its roots are 1 and 2, we get $(3) = (3, \sqrt{7} - 1) \cdot (3, \sqrt{7} - 2)$.
  - For (5) we consider $x^2 - 7$ modulo 5: since it has no roots, we see that (5) remains prime in $\mathbb{Z}[\sqrt{7}]$.
  - For (7) we consider $x^2 - 7$ modulo 7: since it has a repeated root 0, we see $(7) = (7, \sqrt{7})^2 = (\sqrt{7})^2$.

- To finish our discussion here, we will note that almost all of our analysis of the quadratic integer rings $\mathcal{O}_D$ can be extended to general rings of integers of algebraic number fields, as pioneered by Kummer, Dedekind, and Noether in their original development of the theory of rings and modules as applied to number theory.
  - Explicitly, an algebraic number is a complex number that satisfies a polynomial with rational coefficients (such as $i/2$, $\sqrt[3]{2}$, and the roots of $x^5 - x - 1 = 0$), while an algebraic integer is an algebraic number that satisfies a monic polynomial with integer coefficients (such as i and $\sqrt[3]{2}$, but not $i/2$).
  - An algebraic number field is a subfield of $\mathbb{C}$ that is a finite-dimensional vector space over $\mathbb{Q}$ (examples include $\mathbb{Q}(\sqrt{D})$ and $\mathbb{Q}(\sqrt[3]{2})$); all its elements are algebraic numbers.
  - It can be shown that the set of algebraic integers in an algebraic number field K is a subring of K, which is called the ring of integers of the number field. (For example, the ring of integers of $\mathbb{Q}(\sqrt{D})$ is $\mathcal{O}_D$.)
  - Essentially all of the results we have proven then carry over to general rings of integers: ideal divisibility is equivalent to containment, nonzero prime ideals are maximal, nonzero ideals factor as a unique product of prime ideals, and nonzero prime ideals are precisely the ideal factors of (p).
  - In number-theoretic language, if a prime ideal (p) remains prime in a ring of integers, we say (p) is inert. If (p) factors as a product of distinct prime ideals, we say (p) splits, while if (p) has repeated prime factors, we say that p ramifies. The question of when primes split, remain inert, or ramify is a fundamental object of study in algebraic number theory.

8.3 Applications of Factorization In Quadratic Integer Rings

- In this section, we discuss some applications of unique factorization in the quadratic rings $\mathcal{O}_D$.

8.3.1 Factorization in $\mathbb{Z}[i]$ and Sums of Two Squares

- We first discuss factorization in $\mathbb{Z}[i]$, which we have already shown to be a Euclidean domain, a principal ideal domain, and a unique factorization domain.
  - We need only analyze the factorization of primes p, which is fully determined by the ideal factorization of (p) inside $\mathbb{Z}[i]$.
  - Because $N(a+bi) = a^2+b^2$, factorization in $\mathbb{Z}[i]$ is closely related to the question of writing an integer as the sum of two squares, and so by analyzing prime factorizations in $\mathbb{Z}[i]$, we can classify the integers that can be written as the sum of two squares.
- Our first task is to write down the irreducible elements in $\mathbb{Z}[i]$:
- Theorem (Irreducibles in $\mathbb{Z}[i]$): Up to associates, the irreducible elements in $\mathbb{Z}[i]$ are as follows:
  1. The element $1+i$ (of norm 2).
  2. The primes $p \in \mathbb{Z}$ congruent to 3 modulo 4 (of norm p^2).
  3. The distinct irreducible factors $a+bi$ and $a-bi$ (each of norm p) of $p = a^2+b^2$ where $p \in \mathbb{Z}$ is congruent to 1 modulo 4.
  - There are various ways to prove this result using modular arithmetic, but we can establish this result directly from our theorem on factoring the ideal (p).
  - Proof: Since $\mathbb{Z}[i]$ is a Euclidean domain, the irreducible (equivalently, prime) elements in $\mathbb{Z}[i]$ are the generators of its nonzero prime ideals, and these are the ideal factors of the ideals (p) for integer primes p.
  - To find the factorization of (p) in $\mathbb{Z}[i]$, we write down the minimal polynomial $q(x) = x^2+1$ of $\omega = i$ and then determine its factorization modulo p.
  - For $p = 2$ we have $x^2+1 \equiv (x-1)^2 \bmod 2$, so we get the ideal factorization $(2) = (2, i+1)^2$, yielding the element factorization $2 = -(i+1)^2$.
  - For $p \equiv 3 \bmod 4$, the polynomial x^2+1 is irreducible modulo p: we have $\left(\frac{-1}{p}\right) \equiv (-1)^{(p-1)/2} \equiv -1 \pmod{p}$ by Euler's criterion, so -1 is not a square mod p. Thus, (p) is prime in $\mathbb{Z}[i]$, so the element p is irreducible and its norm is p^2.
  - With $p \equiv 1 \bmod 4$, the polynomial x^2+1 factors modulo p because $\left(\frac{-1}{p}\right) \equiv (-1)^{(p-1)/2} \equiv 1 \pmod{p}$ by Euler's criterion, so 1 is a square mod p. With factorization $x^2+1 \equiv (x-r)(x+r) \pmod{p}$, we obtain the ideal factorization $(p) = (p, i-r) \cdot (p, i+r)$.
  - Since $\mathbb{Z}[i]$ is a principal ideal domain, the ideal $(p, i+r) = (a+bi)$ for some a, b that we can compute by applying the Euclidean algorithm to p and $i+r$, and then its conjugate ideal $(p, r-i) = (p, i-r)$ is equal to $(a-bi)$.
  - This yields the ideal factorization $(p) = (a+bi)(a-bi)$ and so we get the element factorization $p = (a+bi)(a-bi)$ up to a unit factor, which by rescaling we may assume is 1. This means $p = (a+bi)(a-bi) = a^2+b^2$, and we have $N(a+bi) = a^2+b^2 = p = N(a-bi)$, so both irreducible factors have norm p as claimed.
- We can now give a method for finding the prime factorization of an arbitrary Gaussian integer:
  - First, find the prime factorization of $N(a+bi) = a^2+b^2$ over the integers $\mathbb{Z}$, and write down a list of all (rational) primes $p \in \mathbb{Z}$ dividing $N(a+bi)$.
  - Second, for each p on the list, find the factorization of p over the Gaussian integers $\mathbb{Z}[i]$.

- Finally, use trial division to determine which of these irreducible elements divide $a+bi$ in $\mathbb{Z}[i]$, and to which powers. (The factorization of $N(a+bi)$ can be used to determine the expected number of powers.)

- Example: Find the factorization of $4+22i$ into irreducibles in $\mathbb{Z}[i]$.
  - We compute $N(4+22i) = 4^2 + 22^2 = 2^2 \cdot 5^3$. The primes dividing $N(4+22i)$ are 2 and 5.
  - Over $\mathbb{Z}[i]$, we find the factorizations $2 = -i(1+i)^2$ and $5 = (2+i)(2-i)$.
  - Now we just do trial division to find the correct powers of each of these elements dividing $4+22i$.
  - Since $N(4+22i) = 2^2 \cdot 5^3$, we should get two copies of $(1+i)$ and three elements from $\{2+i, 2-i\}$.
  - Doing the trial division yields the factorization $4+22i = \boxed{-i \cdot (1+i)^2 \cdot (2+i)^3}$. (Note that in order to have powers of the same irreducible element, we left the unit $-i$ in front of the factorization.)

- The primes appearing in the example above were small enough to factor over $\mathbb{Z}[i]$ by inspection, but if $p \equiv 1 \pmod 4$ is large then it is not so obvious how to factor p in $\mathbb{Z}[i]$. We briefly explain how to find this expression algorithmically.
  - We have the ideal factorization $(p) = (p, i+r) \cdot (p, i-r)$ and then use the Euclidean algorithm to write $(p, i+r) = (a+bi)$. Thus, all we need to do is find a root r of the polynomial $x^2+1 \pmod p$, which is equivalent to finding a square root of -1 modulo p.
  - We can do this using Euler's criterion: for any quadratic nonresidue u modulo p, Euler's criterion tells us that $u^{(p-1)/2} \equiv -1 \pmod p$, and so $u^{(p-1)/4}$ will be a square root of -1.
  - There is no general formula for identifying a quadratic nonresidue modulo an arbitrary prime p, but we can just test small residue classes (or random residue classes) using quadratic reciprocity until we find one.
  - Then, as noted above, to compute the solution to $p = a^2 + b^2$ we can use the Euclidean algorithm in $\mathbb{Z}[i]$ to find a greatest common divisor of p and $r+i$ in $\mathbb{Z}[i]$: the result will be an element $\pi = a+bi$ with $a^2+b^2 = p$.

- Example: Express the prime $p = 3329$ as the sum of two squares.
  - By quadratic reciprocity we have $\left(\frac{3}{3329}\right) = \left(\frac{3329}{3}\right) = \left(\frac{2}{3}\right) = -1$ so 3 is a quadratic nonresidue modulo 3329. Then by successive squaring we may compute $3^{(p-1)/4} \equiv 1729 \pmod p$, meaning that 1729 is a square root of -1 modulo p. (Indeed, $1729^2 + 1 = 898 \cdot 3329$.)
  - Now we compute the gcd of $1729+i$ and 3329 in $\mathbb{Z}[i]$ using the Euclidean algorithm:

$$\begin{aligned} 3329 &= 2(1729+i) + (-129-2i) \\ 1729+i &= -13(-129-2i) + (52-25i) \\ -129-2i &= (-2-i)(52-25i) \end{aligned}$$

  - The last nonzero remainder is $52-25i$, and indeed we can see that $3329 = \boxed{52^2 + 25^2}$.

- As a corollary to our characterization of the irreducible elements in $\mathbb{Z}[i]$, we can deduce the following theorem of Fermat on when an integer is the sum of two squares:

- Theorem (Fermat): Let n be a positive integer, and write $n = 2^k p_1^{n_1} \cdots p_k^{n_k} q_1^{m_1} \cdots q_d^{m_d}$, where $p_1, \cdots, p_k$ are distinct primes congruent to 1 modulo 4 and $q_1, \cdots, q_d$ are distinct primes congruent to 3 modulo 4. Then n can be written as a sum of two squares in $\mathbb{Z}$ if and only if all the m_i are even. Furthermore, in this case, the number of ordered pairs of integers (A, B) such that $n = A^2 + B^2$ is equal to $4(n_1+1)(n_2+1)\cdots(n_k+1)$.
  - Proof: Observe that the question of whether n can be written as the sum of two squares $n = A^2 + B^2$ is equivalent to the question of whether n is the norm of a Gaussian integer $A+Bi$.
  - Write $A+Bi = \rho_1 \rho_2 \cdots \rho_r$ as a product of irreducibles (unique up to units), and take norms to obtain $n = N(\rho_1) \cdot N(\rho_2) \cdot \cdots \cdot N(\rho_r)$.

- By the classification above, if ρ is irreducible in $\mathbb{Z}[i]$, then $N(\rho)$ is either 2, a prime congruent to 1 modulo 4, or the square of a prime congruent to 3 modulo 4. Hence there exists such a choice of ρ_i with $n = \prod N(\rho_i)$ if and only if all the m_i are even.
- Furthermore, since the factorization of $A + Bi$ is unique, to find the number of possible pairs (A, B), we need only count the number of ways to select terms for $A + Bi$ and $A - Bi$ from the factorization of n over $\mathbb{Z}[i]$, which is $n = i^{-k}(1+i)^{2k}(\pi_1\overline{\pi_1})^{n_1} \cdots (\pi_k\overline{\pi_k})^{n_k} q_1^{m_1} \cdots q_d^{m_d}$.
- Up to associates, we must choose $A+Bi = (1+i)^k(\pi_1^{a_1}\overline{\pi_1}^{b_1}) \cdots (\pi_k^{a_k}\overline{\pi_k}^{b_k}) q_1^{m_1/2} \cdots q_d^{m_d/2}$, where $a_i+b_i = n_i$ for each $1 \leq i \leq k$.
- Since there are $n_i + 1$ ways to choose the pair (a_i, b_i), and 4 ways to multiply $A + Bi$ by a unit, the total number of ways is $4(n_1 + 1) \cdots (n_k + 1)$, as claimed.

- Example: Find all ways of writing $n = 6649$ as the sum of two squares.
  - We factor $6649 = 61 \cdot 109$. This is the product of two primes each congruent to 1 modulo 4, so it can be written as the sum of two squares in 16 different ways.
  - We compute $61 = 5^2 + 6^2$ and $109 = 10^2 + 3^2$ (either by the algorithm above or by inspection), so the 16 ways can be found from the different ways of choosing one of $5 \pm 6i$ and multiplying it with $10 \pm 3i$.
  - Explicitly: $(5 + 6i)(10 + 3i) = 32 + 75i$, and $(5 + 6i)(10 - 3i) = 68 + 45i$, so we obtain the sixteen ways of writing 6649 as the sum of two squares as $(\pm 32)^2 + (\pm 75)^2$, $(\pm 68)^2 + (\pm 45)^2$, and the eight other decompositions with the terms interchanged.

8.3.2 Factorization in $\mathcal{O}_{\sqrt{-2}}$ and $\mathcal{O}_{\sqrt{-3}}$

- We can use a similar approach to the one we used in $\mathbb{Z}[i]$ to study factorization in $\mathcal{O}_{\sqrt{-2}} = \mathbb{Z}[\sqrt{2}]$ and $\mathcal{O}_{\sqrt{-3}} = \mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$, which in turn allows us to characterize the integers that can be written in the form $a^2 + 2b^2$ and $a^2 + 3b^2$.
  - As we noted earlier, by using a similar proof to the one we used for $\mathbb{Z}[i]$, we can establish that both $\mathcal{O}_{\sqrt{-2}}$ and $\mathcal{O}_{\sqrt{-3}}$ are Euclidean domains, hence also principal ideal domains and unique factorization domains.
  - We will note that the units in $\mathcal{O}_{\sqrt{-2}}$ are simply ± 1, while the units in $\mathcal{O}_{\sqrt{-3}}$ are the sixth roots of unity: namely, the elements $\frac{\pm 1 \pm \sqrt{-3}}{2}$ and ± 1.
- Our first task is to write down the irreducible elements in these two quadratic integer rings:
- Theorem (Irreducibles in $\mathcal{O}_{\sqrt{-2}}$): Up to associates, the irreducible elements in $\mathcal{O}_{\sqrt{-2}}$ are as follows:
  1. The element $\sqrt{-2}$ (of norm 2).
  2. The primes $p \in \mathbb{Z}$ congruent to 5 or 7 modulo 8 (of norm p^2).
  3. The distinct irreducible factors $a + b\sqrt{-2}$ and $a - b\sqrt{-2}$ (each of norm p) of $p = a^2 + 2b^2$ where $p \in \mathbb{Z}$ is congruent to 1 or 3 modulo 8.

  - Proof: Since $\mathcal{O}_{\sqrt{-2}}$ is a Euclidean domain, the irreducible (equivalently, prime) elements in $\mathcal{O}_{\sqrt{-2}}$ are the generators of its nonzero prime ideals, and these are the ideal factors of the ideals (p) for integer primes p.
  - To find the factorization of (p) in $\mathcal{O}_{\sqrt{-2}}$, we write down the minimal polynomial $q(x) = x^2+2$ of $\omega = \sqrt{-2}$ and then determine its factorization modulo p.
  - For $p = 2$ we have $x^2 + 2 \equiv x^2 \bmod 2$, so we get the ideal factorization $(2) = (\omega)^2$, yielding the element factorization $2 = -(\sqrt{-2})^2$.
  - For $p \equiv 5$ or 7 mod 8, the polynomial x^2+2 is irreducible modulo p: from one of the "secondary" relations from quadratic reciprocity, we know that -2 is a square modulo p if and only if p is congruent to 1 or 3 mod 8. Thus, for $p \equiv 5$ or 7 mod 8, the ideal (p) is prime, so the element p is also prime.
  - With $p \equiv 1$ or 3 mod 8, the polynomial x^2+2 factors modulo p. If the factorization is $x^2+2 \equiv (x-r)(x+r) \pmod{p}$, we obtain the ideal factorization $(p) = (p, \sqrt{-2} - r) \cdot (p, \sqrt{-2} + r)$.

- Since $\mathbb{Z}[\sqrt{-2}]$ is a principal ideal domain, the ideal $(p, \sqrt{-2}+r) = (a+b\sqrt{-2})$ for some a, b that we can compute by applying the Euclidean algorithm to p and $\sqrt{-2}+r$, and then its conjugate ideal $(p, r-\sqrt{-2}) = (p, \sqrt{-2}-r)$ is equal to $(a-b\sqrt{-2})$.
- This yields the ideal factorization $(p) = (a+b\sqrt{-2})(a-b\sqrt{-2})$ and so we get the element factorization $p = (a+b\sqrt{-2})(a-b\sqrt{-2})$ up to a unit factor, which by rescaling we may assume is 1. This means $p = (a+b\sqrt{-2})(a-b\sqrt{-2}) = a^2+2b^2$, and we have $N(a+b\sqrt{-2}) = a^2+2b^2 = p = N(a-b\sqrt{-2})$, so both irreducible factors have norm p as claimed.

- Theorem (Irreducibles in $\mathcal{O}_{\sqrt{-3}}$): Up to associates, the irreducible elements in $\mathcal{O}_{\sqrt{-3}}$ are as follows:

1. The element $\sqrt{-3}$ (of norm 3).
2. The primes $p \in \mathbb{Z}$ congruent to 2 modulo 3 (of norm p^2).
3. The distinct irreducible factors $a+b\sqrt{-3}$ and $a-b\sqrt{-3}$ (each of norm p) of $p = a^2+3b^2$ where $p \in \mathbb{Z}$ is congruent to 1 modulo 3.

- Proof: Since $\mathcal{O}_{\sqrt{-3}}$ is a Euclidean domain, the irreducible (equivalently, prime) elements in $\mathcal{O}_{\sqrt{-3}}$ are the generators of its nonzero prime ideals, and these are the ideal factors of the ideals (p) for integer primes p.
- To find the factorization of (p) in $\mathcal{O}_{\sqrt{-3}}$, we write down the minimal polynomial $q(x) = x^2-x+1$ of $\omega = \dfrac{1+\sqrt{-3}}{2}$ and then determine its factorization modulo p.
- For $p = 3$, we have $x^2-x+1 \equiv (x-2)^2 \pmod{p}$, so we obtain the ideal factorization $(3) = (\omega-2)^2 = (\sqrt{-3})^2$, yielding the element factorization $3 = -(\sqrt{-3})^2$.
- For $p \equiv 2 \bmod 3$, the polynomial x^2-x+1 is irreducible modulo p. For $p = 2$ this can be checked directly, and for odd p, by quadratic reciprocity we have $\left(\dfrac{-3}{p}\right) = \left(\dfrac{-1}{p}\right)\left(\dfrac{3}{p}\right) = (-1)^{(p-1)/2}\left(\dfrac{p}{3}\right)(-1)^{-(p-1)/2} = \left(\dfrac{p}{3}\right)$. When $p \equiv 2 \bmod 3$, this last Legendre symbol is -1, and so -3 is not a square modulo p. Since the roots of x^2-x+1 are $\dfrac{1\pm\sqrt{-3}}{2}$, this means x^2-x+1 has no roots hence is irreducible modulo p. Thus, the ideal (p) is prime, as is the element p.
- For $p \equiv 1 \bmod 3$, we compute instead $\left(\dfrac{-3}{p}\right) = \left(\dfrac{p}{3}\right) = 1$ and so -3 is a square modulo p. Since $2 \neq 0 \bmod p$, this means x^2-x+1 factors modulo p. If the factorization is $x^2-x+1 \equiv (x-r)(x-1+r) \pmod{p}$, we obtain the ideal factorization $(p) = (p, \omega-r)\cdot(p, \omega-1+r)$.
- Since $\mathcal{O}_{\sqrt{-3}}$ is a principal ideal domain, the ideal $(p, \omega-r) = (a+b\sqrt{-3})$ for some a, b that we can compute by applying the Euclidean algorithm to p and $\omega-r$, and then its conjugate ideal will be $(p, \omega-1+r) = (a-b\sqrt{-3})$.
- This yields the ideal factorization $(p) = (a+b\sqrt{-3})(a-b\sqrt{-3})$ and so we get the element factorization $p = (a+b\sqrt{-3})(a-b\sqrt{-3})$ up to a unit factor, which by rescaling we may assume is 1. This means $p = (a+b\sqrt{-3})(a-b\sqrt{-3}) = a^2+3b^2$, and we have $N(a+b\sqrt{-3}) = a^2+3b^2 = p = N(a-b\sqrt{-3})$, so both irreducible factors have norm p as claimed.
- As a final remark, we note that a and b are not necessarily integers, but if both are half-integers, then we can rescale by a unit factor of $\omega\cdot\omega^{-1}$ to make them integers.

- In both of these rings, we can use the same general procedure as in $\mathbb{Z}[i]$ to compute element factorizations.
- First, find the prime factorization of $N(a+b\sqrt{-D}) = a^2+Db^2$ over the integers $\mathbb{Z}$, and write down a list of all (rational) primes $p \in \mathbb{Z}$ dividing $N(a+b\sqrt{-D})$.
- Second, for each p on the list, find the factorization of p in the ring $\mathcal{O}_{\sqrt{-D}}$, which we can do by referring to the lists above, and then solving $p = a^2+Db^2$ in integers a, b whenever this equation has a solution.
- We can find this factorization by inspection for small p, and for large p we can find a solution by solving the quadratic $r^2 \equiv -D \pmod{p}$ and then using the Euclidean algorithm to compute the gcd $a+b\sqrt{-D}$ of p and $\sqrt{-D}+r$ in $\mathcal{O}_{\sqrt{-D}}$.

- Finally, use trial division to determine which of these irreducible elements divide $a + b\sqrt{-D}$ in $\mathcal{O}_{\sqrt{-D}}$ and to which powers. (The factorization of $N(a+b\sqrt{-D})$ can be used to determine the expected number of powers.)

- Example: Find the factorization of $47 + 32\sqrt{-2}$ into irreducibles in $\mathbb{Z}[\sqrt{-2}]$.
  - We compute $N(47+32\sqrt{-2}) = 47^2 + 2 \cdot 32^2 = 3^2 \cdot 11 \cdot 43$, so the primes dividing the norm are 3, 11, and 43.
  - Over $\mathbb{Z}[\sqrt{-2}]$, we find the factorizations $3 = 1^2 + 2 \cdot 1^2 = (1+\sqrt{-2})(1-\sqrt{-2})$, $11 = 3^2 + 2 \cdot 1^2 = (3+\sqrt{-2})(3-\sqrt{-2})$ and $43 = 5^2 + 2 \cdot 3^2 = (5+3\sqrt{-2})(5-3\sqrt{-2})$.
  - Now we just do trial division to find the correct powers of each of these elements dividing $47 + 32\sqrt{-2}$: we will get two of $1 \pm \sqrt{-2}$ and one each of $3 \pm \sqrt{-2}$ and $5 \pm 3\sqrt{-2}$.
  - Doing the trial division yields the factorization $47 + 32\sqrt{-2} = \boxed{(1+\sqrt{-2})^2(3-\sqrt{-2})(5-3\sqrt{-2})}$.

- Example: Find the factorization of $27 - \sqrt{-3}$ into irreducibles in $\mathcal{O}_{\sqrt{-3}}$.
  - We compute $N(27-\sqrt{-3}) = 27^2 + 3 \cdot 1^2 = 2^2 \cdot 3 \cdot 61$, so the primes dividing the norm are 2, 3, and 61.
  - Over $\mathcal{O}_{\sqrt{-3}}$, the element 2 is prime, and we also can find the factorizations $3 = 0 + 3 \cdot 1^2 = -\sqrt{-3}^2$ and $61 = 7^2 + 3 \cdot 2^2 = (7+2\sqrt{-3})(7-2\sqrt{-3})$.
  - Now we just do trial division to find the correct powers of each of these elements dividing $47 + 32\sqrt{-2}$: we get one factor of 2, one factor of $\sqrt{-3}$, and one of $7 \pm 2\sqrt{-3}$.
  - Doing the trial division yields the factorization $27 - \sqrt{-3} = \boxed{\frac{-1-\sqrt{-3}}{2} \cdot 2 \cdot \sqrt{-3} \cdot (7+2\sqrt{-3})}$. Note that the unit factor $\frac{-1-\sqrt{-3}}{2}$ is required to make the product come out correctly. We could, of course, absorb it into any one of the terms, such as by writing instead the factorization $27 - \sqrt{-3} = (-1-\sqrt{-3}) \cdot \sqrt{-3} \cdot (7+2\sqrt{-3})$.

- Using these characterizations of irreducible elements in $\mathbb{Z}[\sqrt{-2}]$ and $\mathcal{O}_{\sqrt{-3}}$, we can describe the integers that can be represented by the two quadratic forms $a^2 + 2b^2$ and $a^2 - ab + b^2$ (or equivalently, $a^2 + 3b^2$):

- Theorem (Integers of the Form $a^2 + 2b^2$): Let n be a positive integer, and write $n = 2^k p_1^{n_1} \cdots p_k^{n_k} q_1^{m_1} \cdots q_d^{m_d}$, where $p_1, \cdots, p_k$ are distinct primes congruent to 1 or 3 modulo 8 and $q_1, \cdots, q_d$ are distinct primes congruent to 5 or 7 modulo 8. Then n can be written in the form $a^2 + 2b^2$ for integers a, b if and only if all the m_i are even. Furthermore, in this case, the number of ordered pairs of integers (A, B) such that $n = A^2 + 2B^2$ is equal to $2(n_1+1)(n_2+1)\cdots(n_k+1)$.
  - Proof: The question of whether n can be written as $n = A^2 + 2B^2$ is equivalent to the question of whether n is the norm of an element $A + B\sqrt{-2} \in \mathbb{Z}[\sqrt{-2}]$.
  - Write $A + B\sqrt{-2} = \rho_1 \rho_2 \cdots \rho_r$ as a product of irreducibles (unique up to units), and take norms to obtain $n = N(\rho_1) \cdot N(\rho_2) \cdot \cdots \cdot N(\rho_r)$.
  - By the classification above, if ρ is irreducible in $\mathbb{Z}[\sqrt{-2}]$, then $N(\rho)$ is either 2, a prime congruent to 1 or 3 modulo 8, or the square of a prime congruent to 5 or 7 modulo 8. Hence there exists such a choice of ρ_i with $n = \prod N(\rho_i)$ if and only if all the m_i are even.
  - Furthermore, since the factorization of $A + B\sqrt{-2}$ is unique, to find the number of possible pairs (A, B), we need only count the number of ways to select terms for $A+B\sqrt{-2}$ and $A-B\sqrt{-2}$ from the factorization of n over $\mathbb{Z}[\sqrt{-2}]$, which is $n = (-1)^k (\sqrt{-2})^{2k} (\pi_1 \overline{\pi_1})^{n_1} \cdots (\pi_k \overline{\pi_k})^{n_k} q_1^{m_1} \cdots q_d^{m_d}$.
  - Up to associates, we must choose $A + B\sqrt{-2} = (\sqrt{-2})^k (\pi_1^{a_1} \overline{\pi_1}^{b_1}) \cdots (\pi_k^{a_k} \overline{\pi_k}^{b_k}) q_1^{m_1/2} \cdots q_d^{m_d/2}$, where $a_i + b_i = n_i$ for each $1 \le i \le k$.
  - Since there are $n_i + 1$ ways to choose the pair (a_i, b_i), and 2 ways to multiply $A + B\sqrt{-2}$ by a unit, the total number of ways is $2(n_1+1)\cdots(n_k+1)$, as claimed.

- Theorem (Integers of the Form a^2+ab+b^2): Let n be a positive integer, and write $n = 3^k p_1^{n_1} \cdots p_k^{n_k} q_1^{m_1} \cdots q_d^{m_d}$, where $p_1, \cdots, p_k$ are distinct primes congruent to 1 modulo 3 and $q_1, \cdots, q_d$ are distinct primes congruent to 2 modulo 3. Then n can be written in the form a^2+ab+b^2 for integers a, b if and only if it can be written in the form a^2+3b^2, if and only if all the m_i are even. Furthermore, in this case, the number of ordered pairs of integers (A, B) such that $n = A^2+AB+B^2$ is equal to $6(n_1+1)(n_2+1)\cdots(n_k+1)$.
  - Proof: The question of whether n can be written as $n = A^2+AB+B^2$ is equivalent to the question of whether n is the norm of an element $A+B\omega \in \mathcal{O}_{\sqrt{-3}}$ where $\omega = \dfrac{1+\sqrt{-3}}{2}$
  - Write $A+B\omega = \rho_1\rho_2\cdots\rho_r$ as a product of irreducibles (unique up to units), and take norms to obtain $n = N(\rho_1)\cdot N(\rho_2)\cdot\cdots\cdot N(\rho_r)$.
  - By the classification above, if ρ is irreducible in $\mathcal{O}_{\sqrt{-3}}$, then $N(\rho)$ is either 3, a prime congruent to 1 modulo 3, or the square of a prime congruent to 2 modulo 3. Hence there exists such a choice of ρ_i with $n = \prod N(\rho_i)$ if and only if all the m_i are even.
  - Furthermore, since the factorization of $A+B\omega$ is unique, to find the number of possible pairs (A, B), we need only count the number of ways to select terms for $A+B\omega$ and $A+B\overline{\omega}$ from the factorization of n over $\mathcal{O}_{\sqrt{-3}}$, which is $n = (-1)^k(\sqrt{-3})^{2k}(\pi_1\overline{\pi_1})^{n_1}\cdots(\pi_k\overline{\pi_k})^{n_k}q_1^{m_1}\cdots q_d^{m_d}$.
  - Up to associates, we must choose $A+B\omega = (\sqrt{-3})^k(\pi_1^{a_1}\overline{\pi_1}^{b_1})\cdots(\pi_k^{a_k}\overline{\pi_k}^{b_k})q_1^{m_1/2}\cdots q_d^{m_d/2}$, where $a_i+b_i = n_i$ for each $1 \le i \le k$.
  - Since there are n_i+1 ways to choose the pair (a_i, b_i), and 6 ways to multiply $A+B\omega$ by a unit, the total number of ways is $6(n_1+1)\cdots(n_k+1)$, as claimed.
  - Finally, for the statement about representations in the form a^2+3b^2, as we have noted, every irreducible element in $\mathcal{O}_{\sqrt{-3}}$ is associate to one in $\mathbb{Z}[\sqrt{-3}]$, so all statements about representability also hold for the norm a^2+3b^2 in this ring.
- Example: Determine whether 21, 101, and 292 can be written in the form a^2+2b^2 and whether they can be written in the form a^2+3b^2 for integers a and b.
  - We have $21 = 3\cdot 7$. Since there is a prime congruent to 7 mod 8 that occurs to an odd power, 21 is not of the form a^2+2b^2. But since all of the primes are either 3 or congruent to 1 modulo 3, 21 is of the form a^2+3b^2.
  - The integer 101 is prime, and it is congruent to 2 modulo 3 and to 5 modulo 8. Therefore, it cannot be written in the form a^2+2b^2 or in the form a^2+3b^2.
  - We have $292 = 2^2\cdot 73$. Since 73 is congruent to 1 modulo 3 and 1 modulo 8, each odd prime is congruent to 1 or 3 modulo 8, so 292 can be written in the form a^2+2b^2. Likewise, since 2 occurs to an even power and 73 is congruent to 1 modulo 3, 292 is also of the form a^2+3b^2.

8.3.3 Some (Additional) Diophantine Equations

- We can exploit unique factorization in various quadratic integer rings to solve Diophantine equations. Here are a few examples of problems of this type:
- Example: Find all integer solutions to the Diophantine equation $x^2+y^2 = z^5$ where x and y are relatively prime.
  - Since squares are 0 or 1 modulo 4, one of x, y must be odd and the other is even, and also z is odd.
  - Now factor the equation inside $\mathbb{Z}[i]$, which as we have shown is a unique factorization domain, as $(x+iy)(x-iy) = z^5$.
  - We now claim that $x+iy$ and $x-iy$ are relatively prime inside $\mathbb{Z}[i]$.
  - To see this, observe that any common divisor must necessarily divide the sum $2x$ and the difference $2iy$, but since x and y are relatively prime integers, this means that the gcd must divide $2 = -i(1+i)^2$.
  - Then the only possible Gaussian prime divisor of the gcd is $1+i$, but $1+i$ does not divide $x+iy$ because x and y have opposite parity.

- Thus, $x+iy$ and $x-iy$ are relatively prime inside $\mathbb{Z}[i]$. Since their product is a fifth power (namely, z^5) and $\mathbb{Z}[i]$ is a UFD, this means that each term must be a fifth power up to a unit factor.
- But since the only units are $\pm 1, \pm i$ and these are all fifth powers (of themselves), we must have $x+iy = (a+bi)^5 = (a^5 - 10a^3b^2 + 5b^4) + (5a^4b - 10a^2b^3 + b^5)i$. Then the conjugate $x-iy$ is $(a-bi)^5$, and $z^5 = (x+iy)(x-iy) = (a^2+b^2)^5$.
- Since all such tuples work, the solutions are of the form $(x,y,z) = (a^5 - 10a^3b^2 + 5b^4,\ 5a^4b - 10a^2b^3 + b^5,\ a^2+b^2)$ for relatively prime integers a and b.

- Example: Show that the only integer solutions to the Diophantine equation $y^2 = x^3 - 2$ are $(3, \pm 5)$.
  - First, observe that y must be odd, for if y were even then we would have $x^3 \equiv 2 \pmod 4$, which is impossible.
  - We can rearrange this equation and then factor it inside $\mathbb{Z}[\sqrt{-2}]$, which as we have shown is a unique factorization domain, as $(y+\sqrt{-2})(y-\sqrt{-2}) = x^3$.
  - We now claim that $y+\sqrt{-2}$ and $y-\sqrt{-2}$ are relatively prime inside $\mathbb{Z}[\sqrt{-2}]$.
  - To see this, observe that any common divisor must necessarily divide their difference $(y+\sqrt{-2}) - (y-\sqrt{-2}) = 2\sqrt{-2} = -(\sqrt{-2})^3$, and since $\sqrt{-2}$ is irreducible in $\mathbb{Z}[\sqrt{-2}]$, the only possible irreducible factor of their difference is $\sqrt{-2}$.
  - But $y+\sqrt{-2}$ cannot be divisible by $\sqrt{-2}$, since this would require y to be even.
  - Thus, $y+\sqrt{-2}$ and $y-\sqrt{-2}$ are relatively prime. Since their product is a cube (namely, x^3) and $\mathbb{Z}[\sqrt{-2}]$ is a UFD, this means that each term must be a cube up to a unit factor.
  - But since the only units are ± 1 and these are both cubes, we must have $y+\sqrt{-2} = (a+b\sqrt{-2})^3 = (a^3 - 6ab^2) + (3a^2b - 2b^3)\sqrt{-2}$, which requires $3a^2b - 2b^3 = 1$.
  - Factoring yields $b(3a^2 - 2b^2) = 1$ and so since a, b are integers, we see that $b = \pm 1$ and then $3a^2 = 2 \pm 1$, which has the two solutions $(a,b) = (\pm 1, -1)$. Then $y = a^3 - 6ab^2 = \pm 5$ and then $x = 3$, and so we obtain the solutions $(3, \pm 5)$ as claimed.
  - Remark: By extending this analysis to the equations of the form $y^2 = x^3 - 2u^6$ for positive integers u with a suitable level of care, we can in fact show that the elliptic curve $y^2 = x^3 - 2$ has rank 1.

- Example: Show that the Diophantine equation $4y^2 = x^3 - 3$ has no integer solutions.
  - First note that y cannot be divisible by 3, since then x would also have to be divisible by 3, but in that case $3 = x^3 - 4y^2$ would be divisible by 9, impossible.
  - Now rearrange the equation and factor it inside the UFD $\mathcal{O}_{\sqrt{-3}}$ as $(2y+\sqrt{-3})(2y-\sqrt{-3}) = x^3$.
  - Any common divisor of $2y+\sqrt{-3}$ and $2y-\sqrt{-3}$ must divide their difference $2\sqrt{-3}$, which is the product of the irreducible elements $\sqrt{-3}$ and 2. Clearly 2 cannot divide $2y+\sqrt{-3}$, and $\sqrt{-3}$ cannot divide it either because y is not divisible by 3.
  - Therefore, $2y+\sqrt{-3}$ and $2y-\sqrt{-3}$ are relatively prime. Since their product is a cube and $\mathcal{O}_{\sqrt{-3}}$ is a UFD, this means that each term must be a cube up to a unit factor.
  - By rescaling and conjugating if necessary, we either have $2y+\sqrt{-3} = (a+b\sqrt{-3})^3$ or $(2y+\sqrt{-3}) \cdot \dfrac{-1+\sqrt{-3}}{2} = (a+b\sqrt{-3})^3$ for some $a, b \in \mathbb{Z}$. However, the second case cannot occur, because the coefficients of the product on the left-hand side are not integers.
  - So we must have $2y+\sqrt{-3} = (a+b\sqrt{-3})^3$, so expanding and comparing coefficients of $\sqrt{-3}$ yields $1 = 3a^2b - 3b^3$, which is impossible since the right-hand side is a multiple of 3.
  - Thus, there are no integer solutions, as claimed.

- We can, with a nontrivial amount of work, also establish the $n = 3$ case of Fermat's conjecture, which was first settled by Euler. For convenience in organizing the proof, we first establish a lemma (which is itself another example of solving a Diophantine equation):

- Lemma (Cubes of the Form $m^2 + 3n^2$): Suppose that m, n are relatively prime integers of opposite parity. If $m^2 + 3n^2 = r^3$, then there exist positive integers a and b with $m = a^3 - 9ab^2$ and $n = 3a^2b - 3b^3$.

- Proof: First observe that if $3|m$ so that $m = 3k$, then we obtain $9k^2 + 3n^2 = r^3$: this forces $3|r$, but then dividing by 3 shows that $n^3 = (r/3)^3 - 3k^2$ so that 3 would also divide n, which is impossible. Thus, $3 \nmid m$.
- Now factor the equation $m^2 + 3n^2 = r^3$ in $\mathcal{O}_{\sqrt{-3}}$ as $(m + n\sqrt{-3})(m - n\sqrt{-3}) = r^3$.
- Any common divisor of $m + n\sqrt{-3}$ and $m - n\sqrt{-3}$ must also divide $2m$ and $2n\sqrt{-3}$, and since m, n are relatively prime, this means the common divisor must divide $2\sqrt{-3}$.
- Since 2 and $\sqrt{-3}$ are irreducible in $\mathcal{O}_{\sqrt{-3}}$, we can see 2 does not divide $m + n\sqrt{-3}$ because m, n have opposite parities, and $\sqrt{-3}$ does not divide $m + n\sqrt{-3}$ because $3 \nmid m$.
- Then since $\mathcal{O}_{\sqrt{-3}}$ is a UFD, we see that $m + n\sqrt{-3}$ must be a unit times a cube: say $m + n\sqrt{-3} = u \cdot (a + b\sqrt{-3})^3$. By negating, conjugating, and replacing $a + b\sqrt{-3}$ with an associate as necessary, we may assume $a, b \in \mathbb{Z}$ and that the unit u is either 1 or $\dfrac{-1 + \sqrt{-3}}{2}$.
- However, if $m + n\sqrt{-3} = \dfrac{-1 + \sqrt{-3}}{2} \cdot (a + b\sqrt{-3})^3$ then since m, n are integers, both a and b must be odd. But then $(-1 + \sqrt{-3})(a + b\sqrt{-3})$ has integer coefficients that are even, as does $(a + b\sqrt{-3})^2$, so the product $m + n\sqrt{-3}$ would have both m and n even, contrary to assumption.
- Therefore, we must have $m + n\sqrt{-3} = (a + b\sqrt{-3})^3 = (a^3 - 9ab^2) + (3a^2b - 3b^3)\sqrt{-3}$ and so $m = a^3 - 9ab^2$ and $n = 3a^2b - 3b^3$, as claimed.

- We can now essentially give Euler's treatment of the $n = 3$ case of Fermat's equation:
- Theorem (Euler): There are no solutions to the Diophantine equation $x^3 + y^3 = z^3$ with $xyz \neq 0$.
  - Proof: Assume $x, y, z \neq 0$ and suppose we have a solution to the equation with $|z|$ minimal.
  - If two of x, y, z are divisible by a prime p then the third must be also, in which case we could divide x, y, z by p and obtain a smaller solution. Thus, without loss of generality, we may assume x, y, z are relatively prime, and so two are odd and the other is even.
  - By rearranging and negating, suppose that x and y are odd and relatively prime. Set $x + y = 2p$ and $x - y = 2q$, so that $x = p + q$ and $y = p - q$, where p, q are necessarily relatively prime of opposite parity.
  - We then obtain a factorization $z^3 = x^3 + y^3 = (x + y)(x^2 - xy + y^2) = 2p \cdot (p^2 + 3q^2)$.
  - First suppose that $3 \nmid p$.
  - Since $p^2 + 3q^2$ is odd, any common divisor of $2p$ and $p^2 + 3q^2$ necessarily divides p and $p^2 + 3q^2$, hence also divides p and $3q^2$. Furthermore, since $3 \nmid p$ this means any common divisor of p and $3q^2$ divides both p and q^2, but these elements are relatively prime.
  - Thus, $2p$ and $p^2 + 3q^2$ are relatively prime, so since their product is a cube, each must be a cube up to a unit factor in $\mathbb{Z}$. But since units are cubes in $\mathbb{Z}$, each term is actually a cube.
  - By the lemma, we then have $p = a^3 - 9ab^2$ and $q = 3a^2b - 3b^3$ for some $a, b \in \mathbb{Z}$, and we also know $2p = 2a(a - 3b)(a + 3b)$ is a cube.
  - We see that $2a$, $a - 3b$, $a + 3b$ must be pairwise relatively prime, since any common divisor would necessarily divide $2a$ and $6b$ hence divide 6, but a cannot be divisible by 3 (since then p, q would both be divisible by 3) and a, b cannot have the same parity (since then both p, q would be even).
  - Therefore, since their product is a cube in $\mathbb{Z}$, each of $2a$, $a - 3b$, and $a + 3b$ must be a cube in $\mathbb{Z}$. But then if $2a = z_1^3$, $a - 3b = x_1^3$, and $a + 3b = y_1^3$, we have $x_1^3 + y_1^3 = z_1^3$, and clearly we also have $0 < |z_1| < |a| < |r| < |z|$. We have therefore found a solution to the equation with a smaller value of z, which is a contradiction.
  - It remains to consider the case where $3|p$, which is quite similar.
  - If we write $p = 3s$ then q, s are relatively prime of opposite parity, and we have $z^3 = 18s \cdot (3s^2 + q^2)$. Since q cannot be divisible by 3 and $3s^2 + q^2$ is odd, any common divisor of $18s$ and $3s^2 + q^2$ must divide s and $3s^2 + q^2$ hence divides s and q^2, but these are relatively prime. Thus $18s$ and $3s^2 + q^2$ are relatively prime, so they are each cubes.

- By the lemma again, we have $q = a^3 - 9ab^2$ and $s = 3a^2b - 3b^3$, where $18s = 3^3 \cdot 2b(a-b)(a+b)$ is a perfect cube. Like before, we see that any common divisor of any pair of $2b$, $a-b$, $a+b$ must divide $2a$ and $2b$ hence divide 2, but a, b must have opposite parity since otherwise q, s would both be even.
- Thus, $2b$, $a-b$, and $a+b$ are all perfect cubes. But then if $a+b = z_1^3$, $a-b = x_1^3$, and $2b = y_1^3$, we have $x_1^3 + y_1^3 = z_1^3$, and clearly we also have $0 < |z_1| = |a+b| < |s| < |z|$. We have again found a solution to the equation with a smaller value of z, which is a contradiction.
- Since we have reached a contradiction in both cases, we are done.

8.3.4 Cubic Reciprocity

- As our next application of our study of the quadratic integer rings, we can develop cubic reciprocity using properties of the ring $\mathcal{O}_{\sqrt{-3}}$.

- Proposition (Arithmetic in $\mathcal{O}_{\sqrt{-3}}$): Let π be a prime of $R = \mathcal{O}_{\sqrt{-3}}$ and let $\omega = \dfrac{-1+\sqrt{-3}}{2}$ denote a nonreal cube root of unity. Then the following are true:

  1. The quotient ring $R/(\pi)$ is a finite field with $N(\pi)$ elements.
     - Proof: If π lies over the prime p, then as we have shown, $\pi | p$. Then (π) contains p, and so there are at most p^2 residue classes modulo π, since any residue class $a + b\omega$ is equivalent to $a' + b'\omega$ where a', b' are the remainders upon dividing a, b by p.
     - Thus, $R/(\pi)$ is a finite ring. Since π is prime, (π) is maximal (since we showed nonzero prime ideals are maximal inside the quadratic integer rings) and so $R/(\pi)$ is in fact a field.
     - Finally, for the statement about the cardinality, if π is associate to $\sqrt{-3}$ then clearly $R/(\pi)$ has 3 residue classes (represented by 0, 1, and 2) and $N(\pi) = 3$.
     - If π is associate to a rational prime $p \equiv 2 \pmod 3$ then $R/(p)$ has p^2 elements (per the calculation above) and $N(\pi) = p^2$.
     - Finally, if π is one of the two conjugate factors of a rational prime $p \equiv 1 \pmod 3$, then $R/(p) \cong R/(\pi) \times R/(\overline{\pi})$ and since both $R/(\pi)$ and $R/(\overline{\pi})$ are fields (and thus have cardinality greater than 1) and $R/(p)$ has cardinality p^2, we must have $\#(R/(\pi)) = \#(R/(\overline{\pi})) = p = N(\pi)$.
  2. For any nonzero residue class α modulo π, we have $\alpha^{N(\pi)-1} \equiv 1 \pmod \pi$.
     - Note that this is a generalization of Euler's theorem for $\mathbb{Z}/m\mathbb{Z}$ to quotients of the quadratic integer ring.
     - Proof: As shown in (1), the quotient ring $R/(\pi)$ is a finite field with $N(\pi)$ elements.
     - The multiplicative group of this finite field then has $N(\pi)-1$ elements. Hence by Lagrange's theorem, any element in this group (i.e., any nonzero residue class) $\alpha^{N(\pi)-1} \equiv 1 \pmod \pi$, as claimed.
  3. If π is not associate to $\sqrt{-3}$, the elements 1, ω, and ω^2 are distinct modulo π, and $N(\pi) - 1$ is divisible by 3.
     - Proof: Suppose that $1 \equiv \omega$, $1 \equiv \omega^2$, or $\omega \equiv \omega^2 \pmod \pi$.
     - Then π necessarily has a nontrivial gcd with $(1-\omega)(1-\omega^2) = 3$, so since π is irreducible, it must be an irreducible factor of 3, hence associate to $\sqrt{-3}$.
     - Taking the contrapositive shows that π is not associate to $\sqrt{-3}$, the elements 1, ω, and ω^2 are distinct modulo π.
     - The second statement then follows by Lagrange's theorem, since $\{1, \omega, \omega^2\}$ is a subgroup of order 3 of the multiplicative group of residues modulo π. Alternatively, we could verify it directly using our characterization of the primes: if π is an integer prime $p \equiv 2 \pmod 3$ then $N(\pi) - 1 = p^2 - 1 \equiv 0 \pmod 3$, and if $\pi\overline{\pi}$ is a prime congruent to 1 modulo 3 then $N(\pi) - 1 = p - 1 \equiv 0 \pmod 3$.

- The idea now is that we can define a cubic residue symbol, which will detect cubes modulo π, in a similar way to how we define the quadratic residue symbol modulo p that detects squares.
  - For the quadratic residue symbol, the idea is to observe that $a^{p-1} - 1 \equiv 0 \pmod p$ by Euler's theorem.

- Thus, when p is odd, we may use the factorization $z^2-1=(z-1)(z+1)$ to factor this expression as $(a^{(p-1)/2}-1)(a^{(p-1)/2}+1)\equiv 0 \pmod p$, which tells us that $a^{(p-1)/2}\equiv 1$ or $-1 \pmod p$.
- Furthermore, the elements with $a^{(p-1)/2}\equiv 1 \pmod p$ will precisely be the squares modulo p: this is Euler's criterion.
- It is easy to see that we can follow an analogous procedure inside $\mathcal{O}_{\sqrt{-3}}/(\pi)$.
- From the proposition above, if π is not associate to $\sqrt{-3}$, then $N(\pi)-1$ is divisible by 3 and $\alpha^{N(\pi)-1}-1\equiv 0 \pmod \pi$.
- Then we may use the factorization $z^3-1=(z-1)(z-\omega)(z-\omega^2)$ to factor the expression as $(\alpha^{(N(\pi)-1)/3}-1)(\alpha^{(N(\pi)-1)/3}-\omega)(\alpha^{(N(\pi)-1)/3}-\omega^2)\equiv 0 \pmod \pi$, and so by unique factorization, this means $\alpha^{(N(\pi)-1)/3}$ is congruent to one of 1, ω, ω^2 modulo π.
- Furthermore (as we will show in a moment) the cubes modulo π are precisely the elements with $\alpha^{(N(\pi)-1)/3}\equiv 1 \pmod \pi$.

- We take this calculation as the definition of our cubic residue symbol:
- Definition: If π is a prime element of $\mathcal{O}_{\sqrt{-3}}$ and $N(\pi)\neq 3$, we define the cubic residue symbol $\left[\frac{\alpha}{\pi}\right]_3\in\{0,1,\omega,\omega^2\}$ to be 0 if $\pi|\alpha$, and otherwise to be the unique value among $\{1,\omega,\omega^2\}$ satisfying $\left[\frac{\alpha}{\pi}\right]_3\equiv\alpha^{(N(\pi)-1)/3} \pmod \pi$.
  - We showed in the proposition that $1,\omega,\omega^2$ are distinct modulo π, and we showed above that $\alpha^{(N(\pi)-1)/3}$ is always congruent to one of $1,\omega,\omega^2$ whenever π does not divide α, so the cubic residue symbol is well-defined.

- The cubic residue symbol also detects cubes, similarly to how the quadratic residue symbol detects squares:
- Proposition (Properties of Cubic Residues): Let π be a prime element of $\mathcal{O}_{\sqrt{-3}}$ with $N(\pi)\neq 3$, and let $\alpha,\beta\in\mathcal{O}_{\sqrt{-3}}$. Then the following hold:

1. If $\alpha\equiv\beta \pmod\pi$ then $\left[\frac{\alpha}{\pi}\right]_3=\left[\frac{\beta}{\pi}\right]_3$.
   - Proof: By definition we have $\left[\frac{\alpha}{\pi}\right]_3\equiv\alpha^{(N(\pi)-1)/3}\equiv\beta^{(N(\pi)-1)/3}\equiv\left[\frac{\beta}{\pi}\right]_3 \pmod\pi$. But since the elements $0,1,\omega,\omega^2$ are distinct modulo π, this congruence actually implies equality: $\left[\frac{\alpha\beta}{\pi}\right]_3=\left[\frac{\alpha}{\pi}\right]_3\left[\frac{\beta}{\pi}\right]_3$.
2. The cubic residue symbol is multiplicative: $\left[\frac{\alpha\beta}{\pi}\right]_3=\left[\frac{\alpha}{\pi}\right]_3\left[\frac{\beta}{\pi}\right]_3$.
   - Proof: By definition we have $\left[\frac{\alpha\beta}{\pi}\right]_3\equiv(\alpha\beta)^{(N(\pi)-1)/3}\equiv\alpha^{(N(\pi)-1)/3}\beta^{(N(\pi)-1)/3}\equiv\left[\frac{\alpha}{\pi}\right]_3\left[\frac{\beta}{\pi}\right]_3 \pmod\pi$, and in the same way as above, this congruence implies equality.
3. We have $\left[\frac{\overline{\alpha}}{\pi}\right]_3=\overline{\left[\frac{\alpha}{\pi}\right]_3}=\left[\frac{\alpha}{\pi}\right]_3^2=\left[\frac{\alpha^2}{\pi}\right]_3$.
   - Proof: For the first equality we have $\left[\frac{\overline{\alpha}}{\pi}\right]_3\equiv\overline{\alpha}^{(N(\pi)-1)/3}\equiv\overline{\alpha^{(N(\pi)-1)/3}}\equiv\overline{\left[\frac{\alpha}{\pi}\right]_3} \pmod\pi$ and again as above this congruence implies equality.
   - For the second equality we note that each of the possible values $0,1,\omega,\omega^2$ has the property that its square equals its complex conjugate.
   - The third equality follows from multiplicativity of the cubic residue symbol.
4. If n is an integer not divisible by π, then $\left[\frac{n}{\pi}\right]_3=1$.

- ◦ Proof: By (3) we have $\overline{\left[\frac{n}{\pi}\right]_3} = \left[\frac{\overline{n}}{\pi}\right]_3 = \left[\frac{n}{\pi}\right]_3$ since n is real. Since $\left[\frac{n}{\pi}\right]_3 \neq 0$ the only possibility is that $\left[\frac{n}{\pi}\right]_3 = 1$.

5. If u is a primitive root modulo π (i.e., an element of order $N(\pi)-1$ modulo π), then $\left[\frac{u}{\pi}\right]_3$ is either ω or ω^2 (i.e., it cannot equal 1).
   - ◦ We remark that a primitive root u must always exist because the multiplicative group of a finite field is always cyclic[3].
   - ◦ Proof: Observe that $\left[\frac{u}{\pi}\right]_3 = u^{(N(\pi)-1)/3}$ cannot be congruent to 1 modulo π since this would mean that the order of u would be at most $(N(\pi)-1)/3$, contradicting the assumption that its order is $N(\pi)-1$.
   - ◦ Thus, since π cannot divide u, $\left[\frac{u}{\pi}\right]_3$ is either ω or ω^2, as claimed.
6. The cubic residue symbol detects cubes: if $\alpha \neq 0 \bmod \pi$, then $\left[\frac{\alpha}{\pi}\right]_3 = 1$ if and only if α is a cubic residue modulo π (which is to say, $\alpha \equiv \beta^3 \pmod{\pi}$ for some β).
   - ◦ Proof: Let u be a primitive root modulo π and write $\alpha = u^k$ for some integer k. Then by (4), since $\left[\frac{\alpha}{\pi}\right]_3 = \left[\frac{u^k}{\pi}\right]_3 = \left[\frac{u}{\pi}\right]_3^k$, and $\left[\frac{u}{\pi}\right]_3$ is either ω or ω^2, we see that that $\left[\frac{\alpha}{\pi}\right]_3 = 1$ if and only if k is a multiple of 3.
   - ◦ But this condition is easily seen to be equivalent to saying that α is a cubic residue: if $\alpha \equiv \beta^3$ then if $\beta = u^r$ we have $\alpha = u^{3r}$, and conversely if k is a multiple of 3 then $\alpha \equiv (u^{k/3})^3$.

- Example: Determine whether $2+\sqrt{-3}$ and $2\sqrt{-3}$ are cubic residues modulo $\pi = 5$ inside $\mathcal{O}_{\sqrt{-3}}$.
  - ◦ Since $N(\pi) = 25$, for $2+\sqrt{-3}$ we must calculate the cubic residue symbol $\left[\frac{2+\sqrt{-3}}{5}\right]_3 \equiv (2+\sqrt{-3})^{(25-1)/3} \equiv (2+\sqrt{-3})^8 \equiv 2+3\sqrt{-3} \pmod{5}$.
  - ◦ Since $\omega = \frac{-1+\sqrt{3}}{2} \equiv 2+3\sqrt{-3} \pmod{5}$, we see $\left[\frac{2+\sqrt{-3}}{5}\right]_3 = \omega$, and so $2+\sqrt{-3}$ is not a cubic residue modulo 5.
  - ◦ For $2\sqrt{-3}$ we calculate $\left[\frac{2\sqrt{-3}}{5}\right]_3 \equiv (2\sqrt{-3})^8 \equiv 1 \pmod{5}$. Thus, $\left[\frac{2\sqrt{-3}}{5}\right]_3 = 1$ and so $2\sqrt{-3}$ is a cubic residue modulo 5.

- In order to handle the situation of associates in $\mathcal{O}_{\sqrt{-3}}$, we select a unique associate for each prime:
- Definition: If π is a prime in $\mathcal{O}_{\sqrt{-3}}$, we say π is primary if $\pi \equiv 2 \pmod{3}$.
  - ◦ If $\pi = a+b\omega$ then this definition is equivalent to saying that $a \equiv 2 \pmod{3}$ and $b \equiv 0 \pmod{3}$.
  - ◦ Example: The primes 2 and $\frac{7+3\sqrt{-3}}{2} = 5+3\omega$ are primary, while $4+\sqrt{-3} = 5+2\omega$ is not primary.
  - ◦ It is straightforward to see that if π is not associate to $\sqrt{-3}$, then exactly one associate of π is primary: explicitly, if $\pi = a+b\omega$ then the associates of π are $\pi = a+b\omega$, $-\pi = (-a)+(-b)\omega$, $\omega\pi = (-b)+(a-b)\omega$, $-\omega\pi = b+(b-a)\omega$, $\omega^2\pi = (b-a)+(-a)\omega$, and $-\omega^2\pi = (a-b)+a\omega$.
  - ◦ It is then straightforward to check that exactly one of b, $a-b$, a is divisible by 3, so two of the associates will have ω-coefficient divisible by 3, and then exactly one will have its coefficient of 1 congruent to 2 modulo 3.

[3] Here is a proof that any finite multiplicative subgroup G of a field F is cyclic: let M be the maximal order among all elements in G; clearly $M \leq \#G$. If g has order M and h is any other element of order k, then if k does not divide M, there is some prime q which occurs to a higher power q^f in the factorization of k than the corresponding power q^e dividing M. Then one may check that $g^{q^f} \cdot h^{k/q^e}$ has order $M \cdot q^{f-e}$, which is impossible because this value is greater than M. Therefore, the order of every element divides M, so the polynomial $p(x) = x^M - 1$ has $\#G$ roots in $F[x]$. But by unique factorization in $F[x]$, this is impossible unless $M \geq \#G$, since a polynomial of degree M can have at most M roots in $F[x]$. Thus, $M = \#G$, so some element has order $\#G$ and G is cyclic.

- We can now state cubic reciprocity in full:
- Theorem (Cubic Reciprocity in $\mathcal{O}_{\sqrt{-3}}$): If π and λ are both primary primes in $\mathcal{O}_{\sqrt{-3}}$ with different norms (i.e., with π, λ both congruent to 2 modulo 3, and with $N(\pi) \neq N(\lambda)$), then $\left[\frac{\pi}{\lambda}\right]_3 = \left[\frac{\lambda}{\pi}\right]_3$.
  - Some aspects of this result were mentioned by Euler and Gauss, and results that are essentially equivalent to this one are implied by some results in Gauss's papers, but the first proof is due to Eisenstein: indeed, the ring $\mathcal{O}_{\sqrt{-3}}$ is occasionally known as the Eisenstein integers for this reason.
  - The proof is relatively involved and is typically broken into three cases: when π and λ are both integer primes, when one is an integer prime, and when both are complex.
  - The first case is trivial, since if p is an integer then $\left[\frac{p}{\lambda}\right]_3 = 1$ regardless of the value of λ, as we showed earlier. The second case requires proving that $\left[\frac{\lambda}{p}\right]_3 = 1$ if p is a prime integer and λ is a prime element, since $\left[\frac{p}{\lambda}\right]_3 = 1$ as noted above. The third case is the most difficult.
- Example: Verify cubic reciprocity for $\pi = \frac{7+3\sqrt{-3}}{2} = 5 + 3\omega$ and $\lambda = 2 + 3\sqrt{-3} = 5 + 6\omega$ in $\mathcal{O}_{\sqrt{-3}}$.
  - We have $N(\pi) = 19$ and $N(\lambda) = 31$.
  - By definition we have $\left[\frac{\lambda}{\pi}\right]_3 \equiv \lambda^{(N(\pi)-1)/3} \equiv (5+6\omega)^6 \equiv \omega^2 \pmod{\pi}$, and we also have $\left[\frac{\pi}{\lambda}\right]_3 \equiv \lambda^{(N(\lambda)-1)/3} \equiv (5+3\omega)^{10} \equiv \omega^2 \pmod{\lambda}$.
  - Thus, we see $\left[\frac{\lambda}{\pi}\right]_3 = \left[\frac{\pi}{\lambda}\right]_3$, precisely as dictated by cubic reciprocity.
- The general approach to most proofs of cubic reciprocity involves manipulation of Gauss sums.
- Definition: A multiplicative character on $\mathbb{F}_p$ is a function $\chi : \mathbb{F}_p^\times \to \mathbb{C}$ such that $\chi(ab) = \chi(a)\chi(b)$ for all $a, b \in \mathbb{F}_p^\times$. If χ is a multiplicative character on $\mathbb{F}_p$, we define the Gauss sum $g_a(\chi) = \sum_{t=1}^{p-1} \chi(t) e^{2\pi i a t/p} \in \mathbb{C}$.
  - We will remark that the values of the Gauss sum $g_a(\chi)$ are the discrete Fourier transform of the function $\chi(t)$, and thus we may convert back and forth between the values of $g_a(\chi)$ and the values $\chi(t)$.
  - For cubic reciprocity, the idea is then to consider the Gauss sums for the cubic character $\chi_\pi(t) = \left[\frac{t}{\pi}\right]_3$ on $\mathbb{F}_p$ where $p = \pi\overline{\pi}$, which encodes all of the information of the cubic residue symbol modulo π.
  - Using the definitions, one may prove various identities involving the Gauss sums for the cubic character χ_π: for example, one can show that $g_a(\chi) = \chi(a)^{-1} g_1(\chi)$, $g_1(\chi_\pi)\overline{g_1(\chi_\pi)} = p$, and $g_1(\chi_\pi)^3 = p\pi$.
- By suitably manipulating these identities, we can then show that $\chi_\lambda(\pi) = \chi_\pi(\lambda)$ for all primary primes λ and π, which establishes cubic reciprocity.
  - We will illustrate by working through the second case of the proof (the third case is more difficult but can be done using a similar method).
  - Proof (Second Case of Cubic Reciprocity): Suppose $q \equiv 2 \pmod 3$ is an integer prime and π is a non-integral prime of $\mathcal{O}_{\sqrt{-3}}$.
  - Take the $(q^2-1)/3$ power of the Gauss-sum identity $g_1(\chi_\pi)^3 = p\pi$ to obtain $g_1(\chi_\pi)^{q^2-1} \equiv (p\pi)^{(q^2-1)/3} \equiv \chi_q(p\pi) = \chi_q(\pi) \pmod q$ because χ_q is multiplicative and $\chi_q(p) = 1$ as noted previously. Thus, $g_1(\chi_\pi)^{q^2} \equiv \chi_q(\pi) g_1(\chi_\pi) \pmod q$.
  - From the definition, we have $g_1(\chi_\pi)^{q^2} = \left[\sum_{t=0}^{p-1} \chi_\pi(t) e^{2\pi i t/p}\right]^{q^2} \equiv \sum_{t=0}^{p-1} \chi_\pi(t)^{q^2} e^{2\pi i q^2 t/p} \pmod q$ since the qth-power map is additive modulo q.
  - Since $q^2 \equiv 1 \pmod 3$ and the value $\chi_\pi(t)$ is zero or a cube root of unity, we have $\chi_\pi(t)^{q^2} = \chi_\pi(t)$ for all t.

- Then we may write $g_1(\chi_\pi)^{q^2} \equiv \sum_{t=0}^{p-1} \chi_\pi(t) e^{2\pi i q^2 t/p} = g_{q^2}(\chi_\pi) = \chi_\pi(q^{-2}) g_1(\chi_\pi) = \chi_\pi(q) g_1(\chi_\pi) \pmod q$ using the Gauss-sum identity $g_a(\chi) = \chi(a)^{-1} g_1(\chi)$ and the fact that $\chi_\pi(q^{-2}) = \chi_\pi(q)$.
- Putting all of this together, we see that $g_1(\chi_\pi)^{q^2}$ is congruent modulo q to both $\chi_q(\pi) g_1(\chi_\pi)$ and to $\chi_\pi(q) g_1(\chi_\pi)$. Multiplying both sides by $\overline{g_1(\chi_\pi)}$ and using the Gauss-sum identity $g_1(\chi_\pi)\overline{g_1(\chi_\pi)} = p$ then yields $\chi_q(\pi) p \equiv \chi_\pi(q) p \pmod q$.
- So, since p is invertible modulo q, we get $\chi_q(\pi) \equiv \chi_\pi(q) \pmod q$, and, at last, this congruence implies the equality $\chi_q(\pi) = \chi_\pi(q)$, which is exactly cubic reciprocity in this case.

- We can use cubic reciprocity to calculate the cubic residue symbol $\left[\frac{\alpha}{\pi}\right]_3$, after we find the prime factorization of the element α, using the same "flip-and-invert" procedure we use for evaluating Legendre symbols.
  - Explicitly, if we write $\alpha = u \cdot (1-\omega)^k \lambda_1 \lambda_2 \cdots \lambda_n$ where the λ_i are primary primes, then we only need to compute the cubic residue symbols $\left[\frac{u}{\pi}\right]_3$, $\left[\frac{1-\omega}{\pi}\right]_3$, and $\left[\frac{\lambda_i}{\pi}\right]_3$.
  - The residue symbol $\left[\frac{u}{\pi}\right]_3$ we can compute using the definition since $u = \pm\omega^k$ and $\left[\frac{\omega}{\pi}\right]_3 = \omega^{(N(\pi)-1)/3}$, so $\left[\frac{\omega}{\pi}\right]_3 = 1$, ω, or ω^2 when $N(\pi) \equiv 1$, 4, or 7 modulo 9 (respectively), and $\left[\frac{-1}{\pi}\right]_3 = 1$.
  - The residue symbol $\left[\frac{1-\omega}{\pi}\right]_3$ is more difficult to compute, but its value can be shown to be equal to $\omega^{2(p+1)/3}$ if $\pi = p$ is an integer prime, and it is equal to $\omega^{2(a+1)/3}$ if $\pi = a + b\omega$ is a primary prime.

8.3.5 Quartic Reciprocity

- We close with a brief discussion of quartic reciprocity, which (like cubic reciprocity) gives a reciprocity law involving fourth powers.
  - The values of the quartic residue symbol will be fourth roots of unity, just as the values of the cubic residue symbol are cube roots of unity, so we will work in the ring $\mathbb{Z}[i]$.

- Proposition (Arithmetic in $\mathbb{Z}[i]$): Let π be a prime of $R = \mathbb{Z}[i]$. Then the following are true:
  1. The quotient ring $R/(\pi)$ is a finite field with $N(\pi)$ elements.
  2. For any nonzero residue class α modulo π, we have $\alpha^{N(\pi)-1} \equiv 1 \pmod \pi$.
  3. If π is not associate to $1+i$, the elements 1, i, -1, and $-i$ are distinct modulo π, and $N(\pi)-1$ is divisible by 4.
     - Proofs: These follow in the same way as the results we showed for $\mathcal{O}_{\sqrt{-3}}$.

- Now we can define the quartic residue symbol.
  - If π is a prime element of odd norm in $\mathbb{Z}[i]$ and $\pi \nmid \alpha$, then since $N(\pi) - 1$ is divisible by 4, we can factor the expression $\alpha^{N(\pi)-1} - 1 \equiv 0$ in $\mathbb{Z}[i]/\pi$ as $(\alpha^{(N(\pi)-1)/4} - 1) \cdot (\alpha^{(N(\pi)-1)/4} + 1) \cdot (\alpha^{(N(\pi)-1)/4} + i) \cdot (\alpha^{(N(\pi)-1)/4} - i) \equiv 0 \pmod \pi$.
  - By unique factorization, this means $\alpha^{(N(\pi)-1)/4}$ is equivalent to one of 1, -1, i, $-i$ modulo π.

- We take this calculation as the definition of our quartic residue symbol:

- Definition: If π is a prime element of $\mathbb{Z}[i]$ and $N(\pi) \neq 2$, we define the quartic residue symbol $\left[\frac{\alpha}{\pi}\right]_4 \in \{0, 1, i, -1, -i\}$ to be 0 if $\pi | \alpha$, and otherwise to be the unique value among $\{1, i, -1, -i\}$ satisfying $\left[\frac{\alpha}{\pi}\right]_4 \equiv \alpha^{(N(\pi)-1)/4} \pmod \pi$.
  - Example: For $\pi = 3$, we have $\left[\frac{1+i}{\pi}\right]_4 \equiv (1+i)^2 \equiv -i \pmod 3$, so $\left[\frac{1+i}{3}\right]_4 = -i$.

- The quartic residue symbol has most of the same properties as the cubic residue symbol:
- Proposition (Properties of Quartic Residues): Let π be a prime element of $\mathbb{Z}[i]$ and $N(\pi) \neq 2$ and let $\alpha, \beta \in \mathbb{Z}[i]$. Then the following hold:
  1. If $\alpha \equiv \beta \pmod{\pi}$ then $\left[\frac{\alpha}{\pi}\right]_4 = \left[\frac{\beta}{\pi}\right]_4$.
  2. The quartic residue symbol is multiplicative: $\left[\frac{\alpha\beta}{\pi}\right]_4 = \left[\frac{\alpha}{\pi}\right]_4 \left[\frac{\beta}{\pi}\right]_4$. Also, $\left[\frac{\overline{\alpha}}{\pi}\right]_4 = \overline{\left[\frac{\alpha}{\pi}\right]_4} = \left[\frac{\alpha}{\pi}\right]_4^3 = \left[\frac{\alpha^3}{\pi}\right]_4$.
  3. If n is an integer not divisible by π, then $\left[\frac{n}{\pi}\right]_4 = 1$ or -1.
  4. If u is a primitive root modulo π (i.e., an element of order $N(\pi)-1$ modulo π), then $\left[\frac{u}{\pi}\right]_4$ is either i or $-i$.
  5. The quartic residue symbol detects fourth powers and squares: if $\alpha \neq 0 \bmod \pi$, then $\left[\frac{\alpha}{\pi}\right]_4 = 1$ if and only if α is a quartic residue modulo π (which is to say, $\alpha \equiv \beta^4 \pmod{\pi}$ for some β), and $\left[\frac{\alpha}{\pi}\right]_4 = -1$ if and only if α is a quadratic residue that is not a quartic residue.
     - Proofs: These follow in the same way as the results we showed for $\mathcal{O}_{\sqrt{-3}}$.
- Example: Find the quartic residues modulo $2+3i$.
  - The nonzero residue classes modulo $\pi = 2+3i$ are represented by the elements 1, 2, 3, ... , 12. The quartic residues are 1, $3 \equiv (2+i)^4$, and $9 \equiv (1+i)^4$. The other 9 classes are quartic nonresidues.
  - We can compute, for example, $\left[\frac{2}{2+3i}\right]_4 \equiv 2^3 \equiv i \pmod{\pi}$, and $\left[\frac{7}{2+3i}\right]_4 \equiv 7^3 \equiv -i \pmod{\pi}$.
- Example: Determine whether $3+3i$, $6-i$, and 6 are quartic residues and whether they are quadratic residues modulo $\pi = 7+2i$ inside $\mathbb{Z}[i]$.
  - Since $N(\pi) = 53$, for $3+3i$ we must calculate the quartic residue symbol $\left[\frac{3+3i}{7+2i}\right]_4 \equiv (3+3i)^{(53-1)/4} \equiv (3+3i)^{13} \equiv -i \pmod{7+2i}$. Thus, $\left[\frac{3+3i}{7+2i}\right]_4 = -i$ and so $3+3i$ is $\boxed{\text{not a quartic or quadratic residue}}$ modulo $7+2i$.
  - For $6-i$ we calculate $\left[\frac{6-i}{7+2i}\right]_4 \equiv (6-i)^{13} \equiv 1 \pmod{7+2i}$. Thus, $\left[\frac{6-i}{7+2i}\right]_4 = 1$ and so $6-i$ $\boxed{\text{is a quartic and quadratic residue}}$ modulo $7+2i$.
  - For 6 we calculate $\left[\frac{6}{7+2i}\right]_4 \equiv (6)^{13} \equiv -1 \pmod{7+2i}$. Thus, $\left[\frac{6}{7+2i}\right]_4 = -1$ and so 6 $\boxed{\text{is not a quartic residue}}$ but $\boxed{\text{is a quadratic residue}}$ modulo $7+2i$.
- We can define a similar notion of a primary prime for $\mathbb{Z}[i]$:
- Definition: A prime element $\pi \in \mathbb{Z}[i]$ is primary if it is congruent to 1 modulo $2+2i$.
  - Example: The primes -3, -7, and $3+2i$ are primary, while 11 and $2+i$ are not.
  - As with the primary elements in $\mathcal{O}_{\sqrt{-3}}$, for all primes except the primes associate to $1+i$ of norm 2, exactly one associate will be primary.
- We can now state quartic reciprocity:
- Theorem (Quartic Reciprocity in $\mathbb{Z}[i]$): If π and λ are distinct primes in $\mathbb{Z}[i]$ congruent to 1 modulo $2+2i$, then $\left[\frac{\pi}{\lambda}\right]_4 = \left[\frac{\lambda}{\pi}\right]_4 \cdot (-1)^{\frac{N(\pi)-1}{4} \cdot \frac{N(\lambda)-1}{4}}$.

- Some aspects of this result (like the other reciprocity laws) were conjectured by Euler, and most of it was known to Gauss; a proof essentially appears in some of his unpublished papers. The first published proof is due to Eisenstein.
- Like with cubic reciprocity, we can establish quartic reciprocity by manipulating the Gauss sums for the quartic character $\chi_\pi(t) = \left[\frac{t}{\pi}\right]_4$.
- The proof is relatively involved and is typically broken into three cases: when π and λ are both integer primes, when one is an integer prime, and when both are complex.
- We will establish the result in one special case, as an illustration, taking as given the Gauss-sum identities $g_a(\chi) = \chi(a)^{-1} g_1(\chi)$, $g_1(\chi_\pi)\overline{g_1(\chi_\pi)} = p$, and $g_1(\chi_\pi)^4 = \pi^3\overline{\pi}$.
- Proof (Second Case): Let q be a prime congruent to 3 modulo 4 (so that $-q$ is the primary element associate to q) and π be a non-integral primary prime with $\pi\overline{\pi} = p$.
- First, taking the $(q+1)/4$th power of the third Gauss-sum identity $g_1(\chi_\pi)^4 = \pi^3\overline{\pi}$ yields $g_1(\chi_\pi)^{q+1} = (\pi^3\overline{\pi})^{(q+1)/4}$.
- Since $\pi^q \equiv \overline{\pi} \pmod{q}$, as can be seen by taking the qth power of $(a+bi)^q$, we see that $g_1(\chi_\pi)^{q+1} \equiv \pi^{(q+1)(q+3)/4} = \pi^{(q^2-1)/4}\pi^{q+1} \equiv \chi_q(\pi)\pi\overline{\pi} \equiv \chi_q(\pi)p \pmod{q}$ by the definition of the quartic residue symbol.
- Also, we have $g_1(\chi_\pi)^q \equiv \left[\sum_{t=1}^{p-1} \chi_\pi(t) e^{2\pi i t/p}\right]^q \equiv \sum_{t=1}^{p-1} \chi_\pi(t)^q e^{2\pi i q t/p} \equiv \sum_{t=1}^{p-1} \overline{\chi_\pi(t)} e^{2\pi i q t/p} \equiv g_q(\overline{\chi_\pi}) \pmod{q}$ because the qth-power map is additive mod q and because $\chi_\pi(t)$ is a fourth root of unity, so since $q \equiv 3 \pmod 4$ the qth power is the same as the complex conjugate.
- But by the first Gauss-sum identity, we have $g_q(\overline{\chi_\pi}) = \overline{\chi_\pi(q)^{-1}} g_1(\overline{\chi_\pi}) = \chi_\pi(-q) g_1(\overline{\chi_\pi})$ since $\chi_\pi(q)$ is a root of unity.
- Putting all of this together yields $\chi_q(\pi)p \equiv g_1(\chi_\pi)^{q+1} \equiv \chi_\pi(-q) g_1(\chi_\pi) g_1(\overline{\chi_\pi}) \equiv \chi_\pi(-q)p \pmod{q}$ using the second Gauss-sum identity. Finally, cancelling the factor of p yields $\chi_\pi(-q) \equiv \chi_q(\pi) \pmod{q}$, and this congruence implies the equality $\chi_\pi(-q) = \chi_q(\pi)$, which is the statement of quartic reciprocity in this case.

- Example: Verify quartic reciprocity for $\pi = 3+2i$ and $\lambda = 5-4i$ in $\mathbb{Z}[i]$.
  - We have $N(\pi) = 13$ and $N(\lambda) = 41$.
  - Then we have $\left[\frac{3+2i}{5-4i}\right]_4 \equiv (3+2i)^{(41-1)/4} \equiv (3+2i)^{10} \equiv i \pmod{5-4i}$, so $\left[\frac{3+2i}{5-4i}\right]_4 = i$.
  - Likewise, $\left[\frac{5-4i}{3+2i}\right]_4 \equiv (5-4i)^{(13-1)/4} \equiv (5-4i)^3 \equiv i \pmod{3+2i}$, so $\left[\frac{5-4i}{3+2i}\right]_4 = i$ as well.
  - Since $\frac{N(\pi)-1}{4} \cdot \frac{N(\lambda)-1}{4}$ is even, the result $\left[\frac{\pi}{\lambda}\right]_4 = \left[\frac{\lambda}{\pi}\right]_4$ is in accordance with quartic reciprocity.
- Example: Verify quartic reciprocity for $\pi = 3+2i$ and $\lambda = 7-2i$ in $\mathbb{Z}[i]$.
  - We have $N(\pi) = 13$ and $N(\lambda) = 53$.
  - Then we have $\left[\frac{3+2i}{7-2i}\right]_4 \equiv (3+2i)^{(53-1)/4} \equiv (3+2i)^{13} \equiv 1 \pmod{7-2i}$, so $\left[\frac{3+2i}{7-2i}\right]_4 = 1$.
  - Likewise, $\left[\frac{7-2i}{3+2i}\right]_4 \equiv (7-2i)^{(13-1)/4} \equiv (7-2i)^3 \equiv -1 \pmod{3+2i}$, so $\left[\frac{7-2i}{3+2i}\right]_4 = -1$.
  - Since $\frac{N(\pi)-1}{4} \cdot \frac{N(\lambda)-1}{4}$ is odd, the result $\left[\frac{\pi}{\lambda}\right]_4 = -\left[\frac{\lambda}{\pi}\right]_4$ is in accordance with quartic reciprocity.

Well, you're at the end of my handout. Hope it was helpful.