Contents

4 Arithmetic and Unique Factorization in Integral Domains

Our goal in this chapter is describe various properties of integral domains related to division algorithms, common divisors, and unique factorization (thereby generalizing many of the properties of $\mathbb{Z}$). We begin by studying Euclidean domains, which are rings that possess a general "division algorithm", and in particular prove that every ideal in a Euclidean domain is principal.

We then enlarge our focus to study general principal ideal domains, in which every ideal is principal, and prove that the elements in principal ideal domains have a unique factorization property. We then broaden our focus again to study the general class of unique factorization domains, and discuss some applications of unique factorization in the classes of rings we have discussed.

Finally, we focus our attention on the quadratic integer rings $\mathcal{O}_D$ and study the properties of ideals 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).

4.1 Euclidean Domains and Principal Ideal Domains

- In this section we will discuss "Euclidean domains", which are integral domains having a division algorithm, and then study the element and ideal structure of these rings. In particular, we will show that every ideal in such a ring is principal.
- Then we will shift our attention to the "principal ideal domains", in which every ideal is principal, and prove that they possess a unique factorization property.

4.1.1 Arithmetic in Integral Domains

- We briefly review some properties of ring 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 the above. Furthermore, $a = ub$ for some unit u clearly implies $a|b$ and $b|a$, and conversely if $a|b$ and $b|a$, then $a = br$
  3. If a and b have a gcd d, then 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.
  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).

- 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" $p = bc$ 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.

- 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.
  - 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.

4.1.2 Euclidean Domains

- Our first 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 $\frac{a+bi}{c+di} = x + iy$ 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 $\frac{r}{c+di} = \frac{a+bi}{c+di} - q = (x-s)+(y-t)i$. Then because $|x - s| \leq \frac{1}{2}$ and $|y - t| \leq \frac{1}{2}$ by construction, the triangle inequality implies $\left|\frac{r}{c+di}\right| \leq \frac{\sqrt{2}}{2}$. Squaring both sides and rearranging yields $N(r) \leq \frac{1}{2}N(c + di)$, as desired.
  - Remark: For the rings $\mathbb{Z}[\sqrt{D}]$ in general, 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 to show that the quadratic integer ring $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is a Euclidean domain for certain small values of D such as $D = -11, -7, -3, -2, 2, 3, 5$.

- 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.

4.1.3 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 integer ring $\mathcal{O}_{\mathbb{Q}(\sqrt{-19})} = \mathbb{Z}[(1 + \sqrt{-19})/2]$.
- Like in Euclidean domains, we can show that any two elements 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.

4.2 Unique Factorization Domains

- We have shown that principal ideal domains possess unique factorization. However, there are rings that possess unique factorization that are not principal ideal domains: for example, in our study of polynomial rings, we have seen that polynomials in $\mathbb{Z}[x]$ can also be factored into products of irreducibles, even though $\mathbb{Z}[x]$ is not a PID.
- This suggests it is worth studying the larger class of integral domains that possess unique factorization.

4.2.1 Arithmetic in Unique Factorization Domains

- 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 $q_1 q_2 \cdots q_d r_1 r_2 \cdots r_k = ab = p$. But since the factorization of ab into irreducibles is unique, we see that p must be associate to one of the irreducibles q_i or r_j.
    - Suppose without loss of generality that $p = q_1 u$: then $q_1 q_2 \cdots q_d r_1 r_2 \cdots r_k = q_1 u$ so upon cancelling q_1 we see that $q_2 \cdots q_d r_1 r_2 \cdots r_k = u$ is a unit, so each term is a unit. Since each of the q_i and r_j is assumed to be irreducible (hence not a unit) we must have $a = q_1$, and so p divides $a = pu^{-1}$, 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.

4.2.2 Unique Factorization in Polynomial Rings

- We would now like to give some additional examples of unique factorization domains beyond the examples of principal ideal domains we have already discussed.

    - As we remarked earlier (though without proof), $\mathbb{Z}[x]$ is a unique factorization domain. Ultimately, this follows from our analysis of factorization in $\mathbb{Q}[x]$, which is a Euclidean domain hence a unique factorization domain: as we have already seen, we can transfer any statement in $\mathbb{Z}[x]$ into one in $\mathbb{Q}[x]$, and more or less vice versa (up to dealing with denominators).
    - More generally, if R is any integral domain, we can try to exploit the factorization properties of polynomials over its field of fractions F. It is natural to ask: for what integral domains R will $R[x]$ be a UFD?
    - Of course, if R itself is not a UFD, then $R[x]$ certainly will not be either, since the factorization of constant polynomials in $R[x]$ reduces immediately to the question of factoring in R.

- By proving a general version of Gauss's lemma, however, we will show that if R is a UFD, then so is $R[x]$.

- Definition: If R is a unique factorization domain, we say a polynomial in $R[x]$ is primitive if the greatest common divisor of its coefficients is equal to 1.

- Proposition (Gauss's Lemma): Let R be a unique factorization domain with field of fractions F, and $p(x) \in R[x]$. If $p(x)$ is reducible in $F[x]$, then $p(x) = f(x)g(x)$ for some $f(x), g(x) \in R[x]$ of positive degree (so in particular, $p(x)$ is reducible in $R[x]$).

  - Proof: First, we observe that, in $F[x]$, any nonzero polynomial $a(x)$ is associate to a primitive polynomial in $R[x]$.
  - To see this, let d be the product of the denominators of $a(x)$: then $d \cdot a(x)$ is a polynomial in $R[x]$. Now let e be the greatest common divisor of the coefficients of $d \cdot a(x)$: then (by the corollary above) we see that $\frac{d}{e} \cdot a(x)$ is a primitive polynomial in $R[x]$; since $\frac{d}{e}$ is a unit in F, this primitive polynomial is associate to $a(x)$.
  - Next, we claim that the product of two primitive polynomials is also primitive.
  - To see this, suppose that $a(x)b(x)$ is not primitive for some $a(x), b(x) \in R[x]$, with $a(x) = a_0 + a_1x + \cdots + a_nx^n$ and $b(x) = b_0 + \cdots + b_mx^m$: then since $a(x)b(x)$ is not primitive, all of its coefficients are divisible by some prime element s.
  - If there is at least one coefficient of each of $a(x)$ and $b(x)$ not divisible by s, suppose that a_i and b_j are the lowest-degree such coefficients. Then the degree-$(i+j)$ term of $a(x)b(x)$ is $a_0b_{i+j} + \cdots + a_{i-1}b_{j+1} + a_ib_j + a_{i+1}b_{j-1} + \cdots + a_{i+j}b_0$, but by hypothesis each term except a_ib_j is divisible by s. This is a contradiction, since this coefficient of $a(x)b(x)$ would then not be divisible by s.
  - Now, returning to the original problem, suppose that $p(x)$ is reducible in $F[x]$ as $p(x) = f_0(x)g_0(x)$ with f_0 and g_0 both of positive degree.
  - By our first observation, both f_0 and g_0 are associate to a primitive polynomial in $R[x]$: say, f, and g respectively.
  - Then (by clearing denominators and cancelling common factors in R) we see that $d \cdot p(x) = e \cdot f(x) \cdot g(x)$ for some relatively prime elements $d, e \in R$.
  - Since d and e are relatively prime, d must divide all coefficients of $f(x)g(x)$, by the corollary above. But $f(x)g(x)$ is primitive by our second observation, so d must be a unit.
  - Then $p(x) = [ed^{-1} \cdot f(x)] \cdot g(x)$ is a nontrivial factorization of $p(x)$ over $R[x]$, as required.

- As suggested by the proof above, the only difference between irreducibility in $F[x]$ and in $R[x]$ is the presence of constant factors from R. More explicitly:

- Corollary: If R is a unique factorization domain with fraction field F, and $p(x) \in R[x]$ is primitive, then $p(x)$ is irreducible in $R[x]$ if and only if $p(x)$ is irreducible in $F[x]$.

  - Proof: By Gauss's lemma, if $p(x)$ is reducible in $F[x]$ then it is reducible in $R[x]$.
  - Conversely, if $p(x) = a(x)b(x)$ is reducible in $R[x]$, then both $a(x)$ and $b(x)$ must be primitive, so neither can be a constant polynomial (as otherwise it would be a unit in $R[x]$). Then $p(x) = a(x)b(x)$ is a nontrivial factorization in $F[x]$.

- We can now finish the proof of our claimed result:

- Theorem (Polynomial Rings and UFDs): If R is an integral domain, then $R[x]$ is a unique factorization domain if and only if R is a unique factorization domain.

  - Proof: If $R[x]$ is a unique factorization domain, then every constant polynomial must factor uniquely into a product of irreducibles, each of which must also be a constant polynomial. But the irreducible constant polynomials of $R[x]$ are precisely the irreducible elements of R, so R must be a UFD.
  - Now suppose that R is a unique factorization domain, and let F be its field of fractions.

- Let $p(x)$ be a nonzero nonunit in $R[x]$, and let d be the greatest common divisor of its coefficients. Then $p(x) = d \cdot q(x)$ where $q(x)$ is a primitive polynomial.
- Since $F[x]$ is a unique factorization domain, we may factor $q(x)$ into a product of irreducibles in $F[x]$. By Gauss's lemma, this yields a factorization of $q(x) = r_1(x)r_2(x)\cdots r_n(x)$, where each polynomial $r_i(x) \in R[x]$ is irreducible in $F[x]$.
- However, since $q(x)$ is primitive, each of the polynomials $r_i(x)$ must also be primitive, so by the corollary above, each polynomial $r_i(x)$ is irreducible in $R[x]$.
- Finally, by factoring d into irreducibles in R, we obtain a factorization $p(x) = [d_1 d_2 \cdots d_m] \cdot [r_1(x)r_2(x)\cdots r_n(x)]$ into irreducible elements in $R[x]$.
- It remains to show that this factorization is unique, so suppose that $p(x)$ has two factorizations into irreducibles and again write $p(x) = d \cdot q(x)$ where $q(x)$ is a primitive polynomial. Since the greatest common divisor d is unique up to associates, and d has a unique factorization into irreducibles by hypothesis, it is enough to show that the factorization of $q(x)$ is unique.
- So suppose $q(x) = r_1(x)r_2(x)\cdots r_n(x) = s_1(x)s_2(x)\cdots s_m(x)$ where the r_i and s_i are irreducible. Again as noted above, each of these polynomials must be primitive, so by the previous corollary they are all irreducible in $F[x]$.
- But since $F[x]$ is a unique factorization domain, we must have $m = n$ and then (by rearranging) that $r_i(x)$ is associate to $s_i(x)$ in $F[x]$ for each $1 \leq i \leq n$.
- Therefore, $r_i(x) = \dfrac{d}{e} s_i(x)$ for some relatively prime elements d and e in R, so clearing denominators yields $er_i(x) = ds_i(x)$: but since $r_i(x)$ and $s_i(x)$ are primitive, we conclude that d and e must be associate in R. If $d = eu$ then we see $r_i(x) = ds_i(x)$, and so r_i is associate to s_i in $R[x]$ for each $1 \leq i \leq n$.
- Thus, we conclude that the factorization in $R[x]$ is unique, so $R[x]$ is a unique factorization domain.

- Corollary: The ring $\mathbb{Z}[x]$ is a unique factorization domain, as is the polynomial ring in two variables $F[x, y] = (F[x])[y]$ for any field F. In particular, there exist unique factorization domains that are not principal ideal domains.
  - Indeed, by a trivial induction, the polynomial ring $F[x_1, x_2, \ldots, x_n]$ in n variables over the field F is also a unique factorization domain, since $F[x_1, x_2, \ldots, x_n]$ is the polynomial ring in the variable x_n with coefficients in $F[x_1, x_2, \ldots, x_{n-1}]$.

4.3 Applications of Unique Factorization

- In this section we collect a few applications of unique factorization in various rings. To treat many of the results we require some preliminary facts about the multiplicative structure of units in rings with 1. We then discuss methods for constructing finite fields and characterize the irreducible elements in the Gaussian integer ring $\mathbb{Z}[i]$ along with some applications to representing integers as sums of two squares.

4.3.1 Orders of Units and Primitive Roots

- We would like to study the behavior of powers of units in a ring with 1.
  - As we have already observed, if u is a unit then so is u^k is also a unit for any integer k, since its inverse is $(u^{-1})^k$.
  - In particular, if there are only finitely many units in R (in particular, if R itself is finite), then the values of the powers of u must eventually repeat.
  - But if $u^a = u^b$ with $a < b$, multiplying both sides by u^{-a} shows that $u^{b-a} = 1$, meaning that some power of u is equal to 1. We give this situation a name:

- Definition: If u is a unit in the ring R, the smallest $k > 0$ such that $u^k = 1$ (if such a k exists) is called the order of u.

- Remark (for those who like group theory): Our use of the word "order" here agrees with the use of the word "order" in group theory, since the set of units in any ring with 1 forms a group under multiplication.
- Example: The powers of 2 in $\mathbb{Z}/11\mathbb{Z}$ are as follows:

$$\begin{array}{cccccccccccc} 2^1 & 2^2 & 2^3 & 2^4 & 2^5 & 2^6 & 2^7 & 2^8 & 2^9 & 2^{10} & 2^{11} & 2^{12} & \cdots \\ 2 & 4 & 8 & 5 & 10 & 9 & 7 & 3 & 6 & 1 & 2 & 4 & \cdots \end{array}$$

  Thus, 2 has order 10 in $\mathbb{Z}/11\mathbb{Z}$.
- Example: The powers of 5 in $\mathbb{Z}/13\mathbb{Z}$ are as follows:

$$\begin{array}{cccccccccc} 5^1 & 5^2 & 5^3 & 5^4 & 5^5 & 5^6 & 5^7 & 5^8 & \cdots \\ 5 & 12 & 8 & 1 & 5 & 12 & 8 & 1 & \cdots \end{array}$$

  Thus, 5 has order 4 in $\mathbb{Z}/13\mathbb{Z}$.
- Example: The powers of 5 in $\mathbb{F}_5[x]/(x^2+1)$ are as follows:

$$\begin{array}{ccccc} x & x^2 & x^3 & x^4 & \cdots \\ x & 4 & 4x & 1 & \cdots \end{array}$$

  Thus, x has order 4 in $\mathbb{F}_5[x]/(x^2+1)$.

- We collect a few useful results about orders.
- Proposition (Properties of Orders): Suppose R is a ring with 1 and u is a unit in R.

1. If $u^n = 1$ for some $n > 0$, then u has finite order and the order of u divides n.
   - Proof: Clearly, if $u^n = 1$ for some $n > 0$, then $u^k = 1$ for some minimal positive integer k by the well-ordering axiom.
   - Now let k be the order of u and apply the division algorithm to write $n = qk + r$ with $0 \leq r < k$: then we have $u^r = u^n(u^k)^{-q} = 1 \cdot 1^{-q} = 1$.
   - If r were not zero, then we would have $u^r = 1$ with $0 < r < k$, which contradicts the definition of order. Thus $r = 0$, meaning that k divides n.
2. If u has order k, then u^n has order $k/\gcd(n,k)$. In particular, if n and k are relatively prime, then u^n also has order k.
   - Proof: Let $d = \gcd(n,k)$: then $(u^n)^{k/d} = (u^k)^{n/d} = 1^{n/d} = 1$, so the order of u^n cannot be larger than k/d.
   - Furthermore, if $1 = (u^n)^a = u^{na}$, the result above implies that k divides na, so that k/d divides $(n/d)a$.
   - But since k/d and n/d are relatively prime, this implies k/d divides a, and so $a \geq k/d$.
   - Thus, the order of u^n is equal to k/d as claimed. The second statement is simply the case $d = 1$.
3. If $u^n = 1$ and $u^{n/p} \neq 1$ for any prime divisor p of n, then u has order n.
   - Proof: Suppose u has order k: then by the above, k must divide n. If $k < n$, then there must be some prime p in the prime factorization of n that appears to a strictly lower power in the factorization of k: then k divides n/p.
   - But then $u^{n/p}$ would be an integral power of $u^k = 1$, so that $u^{n/p} = 1$, which is a contradiction. Thus, $r = n$.
4. If R is commutative, u has order k, and w has order l, where k and l are relatively prime, then uw has order kl.
   - Proof: First observe that $(uw)^{kl} = (u^k)^l(w^l)^k = 1$, uw has some finite order $d \leq kl$.
   - Since $(uw)^d = 1$, raising to the kth power yields $1 = (uw)^{dk} = w^{dk}$, so l divides dk.
   - Then since l and k are relatively prime, this implies l divides d. By a symmetric argument, k divides d. Since l and k are relatively prime, we see kl divides d, and so the only possibility is $d = kl$.

- In the case of a commutative ring with finitely many units, we can say more about the possible order of a unit:

- Theorem (Euler): If R is a commutative ring with 1 with a finite number n of units, then the order of any unit in R divides n. In particular, if u is relatively prime to m, then $u^{\varphi(m)} \equiv 1 \pmod{m}$, where $\varphi(m)$ denotes Euler's φ-function.
    - Remark: This result actually holds even when R is a noncommutative ring, but the proof is more difficult (it follows from Lagrange's theorem in group theory). Since we will not need the general version, we will give the proof in the commutative case only.
    - Proof: Suppose u is a unit in R, and let the set of all units in R be $w_1, w_2, \ldots, w_n$.
    - Consider the elements $uw_1, uw_2, \ldots, uw_n$: there n elements in this list, they are all units, and they are all distinct since $uw_1 = uw_2$ implies $w_1 = w_2$. Thus, they must simply be the elements $w_1, w_2, \ldots, w_n$ again (possibly in a different order).
    - Thus, multiplying all the elements together yields $(uw_1)(uw_2)\cdots(uw_n) = w_1 w_2 \cdots w_n$, whence $u^n(w_1 w_2 \cdots w_n) = w_1 w_2 \cdots w_n$. Then cancelling the unit factor $w_1 w_2 \cdots w_n$ yields $u^n = 1$, and so the order of u divides n.
    - The second statement follows by taking $R = \mathbb{Z}/m\mathbb{Z}$, since $\varphi(m)$ is the number of units in $\mathbb{Z}/m\mathbb{Z}$.
- As a corollary, we obtain a related result that (in the special case of $\mathbb{Z}/p\mathbb{Z}$) is known as Fermat's little theorem:
- Corollary (Fermat): If F is a finite field, then $a^{|F|} = a$ for any $a \in F$. In particular, $a^p \equiv a \pmod{p}$ for every prime p.
    - Proof: If $a = 0$ then clearly $a^{|F|} = a$. If $a \neq 0$ then since F is a field we see that a is a unit: then by Euler's theorem we have $a^{|F|-1} = 1$ so that $a^{|F|} = a$.
    - The second statement follows by taking $F = \mathbb{Z}/p\mathbb{Z}$.
- By Euler's theorem, the order of any unit in R divides the number of units in R. The case when equality occurs is sufficiently useful that we give it a name:
- Definition: If R is a commutative ring with 1 with n units, and the unit u has order n, then we say that u is a primitive root in R.
    - Example: In $R = \mathbb{Z}/5\mathbb{Z}$, the powers of 2 are 2, 4, 3, 1, so 2 is a primitive root in $\mathbb{Z}/5\mathbb{Z}$ since it has order 4.
    - Example: In $R = \mathbb{F}_3[x]/(x^2+1)$, the powers of $x+1$ are $x+1$, $2x$, $2x+1$, 2, $2x+2$, x, $x+2$, 1, so $x+1$ is a primitive root in R since it has order 8 (there are 8 units in R because R is a field).
    - Example: In $R = \mathbb{Z}/9\mathbb{Z}$, the powers of 2 are 2, 4, 8, 7, 5, and 1, so 2 is a primitive root in $\mathbb{Z}/9\mathbb{Z}$ since it has order 6 (and there are 6 units in $\mathbb{Z}/9\mathbb{Z}$).
    - We remark that primitive roots may not necessarily exist.
    - Non-Example: There is no primitive root in $\mathbb{Z}/15\mathbb{Z}$: the units are 1 (order 1), 2 (order 4), 4 (order 2), 7 (order 4), 8 (order 4), 11 (order 2), and 14 (order 2), and none of these is a primitive root.
    - Remark (for those who like group theory): R has a primitive root u if and only if the group of units $R^\times$ is a cyclic group generated by u.
- In the two examples above where R was a finite field, R possessed a primitive root. This is true in general:
- Theorem (Primitive Roots in Finite Fields): If F is a finite field, then F has a primitive root.
    - Our proof is nonconstructive: we will establish the existence of a primitive root without explicitly finding one.
    - Proof: First we will show that if M is the maximal order among all units in F, then the order of every unit divides M. Then we will show that $M = |F| - 1$, which will establish the existence of a primitive root in F.
    - For the first claim, suppose u has order M, and let w be any other unit of order k. If k does not divide M, then 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.

- Observe that the element u^{q^f} has order M/q^f, and the element w^{k/q^e} has order q^e. Since these two orders are relatively prime and F is commutative, by our results about orders we see that the element $u^{q^f} \cdot w^{k/q^e}$ has order $M \cdot q^{f-e}$. This is a contradiction because this element's order is larger than M. Thus, k divides m as claimed.
- For the second claim, since M is the maximal order of among all units of F, then by Euler's theorem we know that $M \leq |F| - 1$.
- Furthermore, by the above, we know that all units in F then have order dividing M, so the polynomial $p(x) = x^M - 1$ has $|F|-1$ roots in $F[x]$. But since $F[x]$ is a unique factorization domain, this is impossible unless $M \geq |F| - 1$, since a polynomial of degree M can only have at most M roots in $F[x]$.
- Hence we conclude $M = |F| - 1$, meaning that some element has order $|F| - 1$: this element is then a primitive root.

4.3.2 Finite Fields and Irreducible Polynomials in $\mathbb{F}_p[x]$

- We have already seen that we can construct finite fields as quotient rings of the form $\mathbb{F}_p[x]/(q)$ where q is an irreducible polynomial in $\mathbb{F}_p[x]$.
  - In particular, if we can establish the existence of an irreducible polynomial in $\mathbb{F}_p[x]$ of degree n, we obtain a construction for a finite field with p^n elements.
  - If q is irreducible of degree n, then by Fermat's theorem in $F = \mathbb{F}_p[x]/(q)$, we see that every element $\overline{a} \in F$ has the property that $\overline{a}^{p^n} = \overline{a}$. In particular, for $a = x$, we see that $\overline{x}^{p^n} = \overline{x}$, meaning that $\overline{x^{p^n} - x} = \overline{0}$: but this is simply another way of saying that $x^{p^n} - x$ is divisible by $q(x)$.
  - We therefore see that irreducible polynomials in $\mathbb{F}_p[x]$ of degree n will appear in the factorization of the polynomial $x^{p^n} - x$ in $\mathbb{F}_p[x]$: this suggests that we may be able to study irreducible polynomials by examining the factorization of $x^{p^n} - x$.
  - Example: For $n = 2$ and $p = 2$, we find the irreducible factorization $x^4 - x = x(x+1)(x^2 + x + 1)$.
  - Example: For $n = 3$ and $p = 2$, we find the irreducible factorization $x^8 - x = x(x+1)(x^3 + x^2 + 1)(x^3 + x + 1)$.
  - Example: For $n = 4$ and $p = 2$, we find the irreducible factorization $x^{16} - x = x(x+1)(x^2 + x + 1)(x^4 + x^3 + 1)(x^4 + x + 1)(x^4 + x^3 + x^2 + x + 1)$.
  - Example: For $n = 2$ and $p = 3$, we find the irreducible factorization $x^9 - x = x(x+1)(x+2)(x^2+2)(x^2 + x + 2)(x^2 + 2x + 2)$.
  - Example: For $n = 2$ and $p = 5$, the irreducible factorization of $x^{25} - x$ is the product of the terms x, $x+1$, $x+2$, $x+3$, $x+4$, x^2+2, x^2+3, x^2+x+1, x^2+x+2, x^2+2x+3, x^2+2x+4, x^2+3x+3, x^2+3x+4, x^2+4x+1, and x^2+4x+2.
  - We notice (especially in the $p = 5$ example) that the irreducible factors all appear to be of small degree, and that there are no repeated factors. More specifically, the factorization seems to consist of the product of all monic irreducible polynomials of degree dividing n.
- Theorem (Factorization of $x^{p^n} - x$ in $\mathbb{F}_p[x]$): For any prime p and any positive integer n, the polynomial $x^{p^n} - x$ factors in $\mathbb{F}_p[x]$ as the product of all monic irreducible polynomials over $\mathbb{F}_p$ of degree dividing n.
  - Proof: Let $q(x) = x^{p^n} - x$ and $R = \mathbb{F}_p[x]$. We prove the result in the following way: first, we show that $q(x)$ has no repeated factors. Second, we show that every irreducible polynomial of degree dividing n divides $q(x)$. Finally, we show that no other irreducible polynomial can divide $q(x)$.
  - For the first part, suppose that $p(x)^2$ divides $q(x)$, so that $q(x) = p(x)^2 r(x)$. Then by differentiating, we have $q'(x) = p(x)[2p'(x)r(x) + p(x)r'(x)]$, so $p(x)$ also divides $q'(x)$.
  - But $q'(x) = p^n x^{p^n - 1} - 1 = -1$, so $p(x)$ must be a constant polynomial. Thus, $q(x)$ cannot have any repeated irreducible factors.
  - Before starting the rest of the proof, we first show that the gcd of $p^n - 1$ and $p^d - 1$ is $p^{\gcd(n,d)} - 1$ for any positive integer d.
    - To see this, write $n = qd + r$, and let $a = p^r(p^{(q-1)d} + p^{(q-2)d} + \cdots + p^d + 1)$.

    - Some arithmetic will show that $p^n - 1 = (p^d - 1)a + (p^r - 1)$.
    - Then $\gcd(p^n - 1, p^d - 1) = \gcd(p^d - 1, p^r - 1)$. But this means we can perform the Euclidean algorithm on the exponents without changing the gcd.
    - The end result is $p^{\gcd(n,d)} - 1$, so this is the desired gcd.
  - For the second part, suppose that $s(x) \in \mathbb{F}_p[x]$ is an irreducible polynomial of degree d, where $n = ad$. If $s(x) = x$ then the result is obvious, so assume $s(x) \neq x$.
  - We know that $\mathbb{F}_p[x]/(s)$ is a finite field having p^d elements, so by Euler's Theorem we see that $x^{p^d - 1} \equiv 1 \pmod{s}$.
  - But, by the lemma, $p^d - 1$ divides $p^n - 1$, so raising to the appropriate power modulo s shows $x^{p^n - 1} \equiv 1 \pmod{s}$. We conclude that s divides $x^{p^n} - x$, as desired.
  - For the final part, suppose $s(x) \in \mathbb{F}_p[x]$ is an irreducible polynomial that divides $x^{p^n} - x$ and has degree d not dividing n. Since $s(x) \neq x$, we can assume s divides $x^{p^n - 1} - 1$.
  - As above, $\mathbb{F}_p[x]/(s)$ is a finite field having p^d elements, so by Euler's Theorem in F, we see that $a^{p^d - 1} \equiv 1 \pmod{s}$ for every nonzero $a \in F$.
  - Since $a^{p^n - 1} \equiv 1 \pmod{s}$ holds for every nonzero $a \in \mathbb{F}_p[x]/(s)$ by the above assumptions, we conclude that $a^{p^{\gcd(d,n)} - 1} \equiv 1 \pmod{s}$.
  - But this is impossible, because $q(t) = t^{p^{\gcd(d,n)} - 1} - 1$ is then a polynomial of degree $p^{\gcd(d,n)} - 1$ which has $p^d - 1$ roots over the field $\mathbb{F}_p$. This completes the final part of the proof, so we are done.

- Corollary: For any prime p and any positive integer n, there exists a finite field having p^n elements.
  - It can be shown that a finite field must have prime-power order, so this result completely characterizes the number of elements that a finite field can have.
  - Proof: By taking degrees, we see that the sum of the degrees of all monic irreducibles of degree dividing d is p^d, so in particular the sum of the degrees of all monic irreducibles of degree exactly d is at most p^d.
  - Thus, the sum of the degrees of all monic irreducibles of degree $\leq n - 1$ is at most $1 + p + \cdots + p^{n-1}$.
  - Since $p^n - (1 + p + \cdots + p^{n-1}) > p^n - (p - 1)(1 + p + \cdots + p^{n-1}) = 1$, we conclude that there is at least 1 monic irreducible polynomial of degree n.

- We can refine the argument above to give an exact count:
  - Let $f_p(n)$ be the number of monic irreducible polynomials of exact degree n in $\mathbb{F}_p[x]$.
  - The theorem says that $p^n = \sum_{d|n} d f_p(d)$, since both sides count the total degree of the product of all irreducible polynomials of degree dividing n. Using this recursion, we can compute the first few values:

n	1	2	3	4	5	6	7	8
$f_p(n)$	p	$\frac{1}{2}(p^2 - p)$	$\frac{1}{3}(p^3 - p)$	$\frac{1}{4}(p^4 - p^2)$	$\frac{1}{5}(p^5 - p)$	$\frac{1}{6}(p^6 - p^3 - p^2 + p)$	$\frac{1}{7}(p^7 - p)$	$\frac{1}{8}(p^8 - p^4)$

  - For example, we see that there are $(3^7 - 3)/7 = 312$ monic irreducible polynomials of degree 7 over $\mathbb{F}_3$.

- In fact, we can use the recursion to write down a general formula (essentially):

- Definition: The Möbius function is defined as $\mu(n) = \begin{cases} 0 & \text{if } n \text{ is divisible by the square of any prime} \\ (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes} \end{cases}$. In particular, $\mu(1) = 1$.

- Proposition (Möbius Inversion): If $f(n)$ is any sequence satisfying a recursive relation of the form $g(n) = \sum_{d|n} f(d)$, for some function $g(n)$, then $f(n) = \sum_{d|n} \mu(d) g(n/d)$.
  - Proof: First, consider the sum $\sum_{d|n} \mu(d)$: we claim it is equal to 1 if $n = 1$ and 0 if $n \neq 0$.

- To see this, if $n = p_1^{a_1} \cdots p_k^{a_k}$, the only terms that will contribute to the sum $\sum_{d|n} \mu(d)$ are those values of $d = p_1^{b_1} \cdots p_k^{b_k}$ where each b_i is 0 or 1. If $k > 0$, then half of these 2^k terms will have $\mu(d) = 1$ and the other half will have $\mu(d) = -1$, so the sum is zero. Otherwise, $k = 0$ means that $n = 1$, in which case the sum is clearly 1.
- Now we prove the desired result by (strong) induction. It clearly holds for $n = 1$, so now suppose the result holds for all $k < n$.
- By hypothesis and induction, $\sum_{d|n} \mu(d)g(n/d) = \sum_{d|n} \mu(d) \sum_{d'|(n/d)} f(d') = \sum_{dd'|n} \mu(d)f(d') = \sum_{d'|n} f(d') \sum_{d|(n/d')} \mu(d)$, but this last sum is simply $f(n)$, because $\sum_{d|(n/d')} \mu(d)$ is zero unless n/d' is equal to 1.

- By applying Möbius inversion to our particular function $f_p(n)$, we immediately obtain the following:
- Corollary: The number of monic irreducible polynomials of degree n in $\mathbb{F}_p[x]$ is $f_p(n) = \frac{1}{n} \sum_{d|n} p^{n/d} \mu(d)$.

- From this corollary, we see that $f_p(n) = \frac{1}{n} p^n + O(p^{n/2})$, where the "big-O" notation means that the error is of size bounded above by a constant times $p^{n/2}$ as $n \to \infty$.
- This has the following interesting reinterpretation: let $X = p^n$ be the number of polynomials in $\mathbb{F}_p[x]$ of degree less than n.
- Now we ask: of these X polynomials, how many of them are prime (i.e., irreducible)?
- This is simply $f_p(n) = \frac{1}{n} p^n + O(p^{n/2}) = \frac{X}{\log_p(X)} + O(\sqrt{X})$.
- In other words: the number of "primes less than X" is equal to $\frac{X}{\log_p(X)}$, up to a bounded error term.
- Notice how very similar this statement is to the statement of the prime number theorem for the integers $\mathbb{Z}$! (This is not a coincidence.)

4.3.3 Factorization in $\mathbb{Z}[i]$

- We would like to analyze prime factorization in the ring $R = \mathbb{Z}[i]$.

- Since $\mathbb{Z}[i]$ is a Euclidean domain, we know that prime elements are the same as irreducible elements, but we will generally use the term "irreducible" element when referring to $\mathbb{Z}[i]$, so as not to cause too much confusion when also we refer to prime numbers in $\mathbb{Z}$.
- We will reserve the letter p for a prime integer (in $\mathbb{Z}$) and we will use π to denote an irreducible element in $\mathbb{Z}[i]$. (The use of the letter π is traditional, and should not cause confusion with the real number π.)
- Recall that in $\mathbb{Z}[i]$, we have the norm map $N(a+bi) = a^2+b^2 = |a+bi|^2$, taking values in the nonnegative integers, that this map satisfies $N(zw) = N(z)N(w)$, and z is a unit in $\mathbb{Z}[i]$ if and only if $N(z) = 1$ (which is to say, $z \in \{1, -1, i, -i\}$).
- We also observe that if $N(z) = p$ is a prime number, then z is irreducible, since any factorization would necessarily contain a term of norm 1 hence be a unit. Thus, for example, we see immediately that the elements $1 + i$ and $2 + i$ are irreducible in $\mathbb{Z}[i]$ since $N(1 + i) = 2$ and $N(2 + i) = 5$.
- However, there are irreducible elements whose norms are not prime. For example, if $3 = zw$ for some nonunits z and w, then the only possibility would be $N(z) = N(w) = 3$, but there are no Gaussian integers of norm 3 since there are no integer solutions to $3 = a^2 + b^2$.
- We can in fact extend this line of reasoning: $\pi \in \mathbb{Z}[i]$ is irreducible, then certainly π divides $N(\pi)$. But since π is a prime element of $\mathbb{Z}[i]$, we see that π must divide one of the (integer) prime factors of the integer $N(\pi)$. Thus, to characterize the irreducible elements of $\mathbb{Z}[i]$, we need to study how primes $p \in \mathbb{Z}$ factor in $\mathbb{Z}[i]$.

- Proposition (Sums of Squares and Primes in $\mathbb{Z}[i]$): If p is a prime integer, then p is irreducible in $\mathbb{Z}[i]$ if and only if p is not the sum of two squares (of integers). In particular, 2 is reducible in $\mathbb{Z}[i]$, while any prime congruent to 3 modulo 4 is irreducible in $\mathbb{Z}[i]$.
  - Proof: Suppose that $p = (a+bi)(c+di)$ for some nonunits $a+bi$ and $c+di$, where p is a prime in $\mathbb{Z}$.
  - Taking norms yields $p^2 = N(p) = (a^2+b^2)(c^2+d^2)$, and now since $a+bi$ and $c+di$ are not units, both a^2+b^2 and c^2+d^2 must be larger than 1.
  - The only possibility is $a^2+b^2 = c^2+d^2 = p$, so we see that $p = a^2+b^2$ for some integers a and b.
  - Conversely, if $p = a^2+b^2$ for some integers a and b, we have the factorization $p = (a+bi)(a-bi)$.
  - For the last statement, clearly $2 = 1^2+1^2$. Furthermore, any square is 0 or 1 modulo 4, so the sum of two squares cannot be 3 modulo 4.
- We are now left to analyze primes congruent to 1 modulo 4.
  - By testing a few small cases like $5 = (2-i)(2+i)$ and $13 = (3+2i)(3-2i)$, it would appear that such primes always factor into a product of two complex-conjugate irreducible factors in $\mathbb{Z}[i]$. This turns out to be the case.
- Proposition (Factorization of 1 mod 4 Primes): If $p \equiv 1 \pmod 4$, then p is a reducible element in the ring $\mathbb{Z}[i]$, and its factorization into irreducibles is $p = (a+bi)(a-bi)$ for some a and b with $a^2+b^2 = p$.
  - Proof: First we will show that there exists some integer n such that p divides n^2+1, and then we use the result to show that p is reducible in $\mathbb{Z}[i]$.
  - For the first part, let $p = 4k+1$ and let u be a primitive root modulo p (which we have shown necessarily exists).
  - Then $u^{4k} \equiv 1 \bmod p$, so $u^{2k} \equiv -1 \pmod p$, since its square is 1 but it cannot equal 1 (as otherwise u would have order $\leq 2k$ and thus not be a primitive root).
  - Then $u^k = n$ is an element whose square is -1 modulo p, so p divides the integer n^2+1.
  - For the second part, we see that p divides $n^2+1 = (n+i)(n-i)$ in $\mathbb{Z}[i]$.
  - Then, since p is a real number, if p divides one of $n \pm i$ then taking complex conjugates would show that p also divides the other. But this is not possible, since then p would divide $(n+i)-(n-i) = 2i$, which it clearly does not.
  - Therefore, hence, p is not a prime element in $\mathbb{Z}[i]$, so it must be reducible. Then by the previous proposition, there exist integers a and b with $p = a^2+b^2$.
  - Then $N(a+bi) = N(a-bi) = p$ so these two elements are both irreducible, meaning that the factorization of p in $\mathbb{Z}[i]$ is $p = (a+bi)(a-bi)$ as claimed.
- This completes our characterization of the irreducible elements in $\mathbb{Z}[i]$. Explicitly:
- 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.

  - Proof: The above propositions show that each of these are irreducible elements; we need only show there are no others. So suppose $\pi = a+bi$ is an irreducible element in $\mathbb{Z}[i]$.
  - Then $N(\pi) = p_1 p_2 \cdots p_k$ for some (integer) primes $p_i \in \mathbb{Z}$; since π is a prime element we conclude that it must divide one of the p_i. But we have characterized how p_i factors into irreducibles in $\mathbb{Z}[i]$, so it must be associate to one of the elements on our list above.
- Using this characterization of irreducible elements, we can describe a method for factoring an arbitrary Gaussian integer into irreducibles. (This is the "prime factorization" in $\mathbb{Z}[i]$.)

- 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 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.
  - Per the proof given above, we first want to find n such that p divides n^2+1, which is equivalent to finding a square root of -1 modulo p.
  - One way to search for such values is to choose a (random) unit u modulo p: then since $u^{p-1} \equiv 1 \pmod{p}$, we know that the square of $u^{(p-1)/2}$ will be $\equiv 1 \pmod{p}$. It can be proven that there is a $1/2$ probability that $u^{(p-1)/2}$ will be congruent to -1 modulo p, and in this case the value $u^{(p-1)/4}$ will be a square root of -1 modulo p. By trying various choices for u, we can eventually find the desired[1] n.
  - Now suppose we have computed such an n: if we factor $p = \pi\overline{\pi}$ in $\mathbb{Z}[i]$, then since π divides $n^2+1 = (n+i)(n-i)$ and π is a prime element, either π divides $n+i$ or π divides $n-i$. Equivalently, either π divides $n+i$ or $\overline{\pi}$ divides $n+i$.
  - Furthermore, since p clearly does not divide $n+i$, we see that exactly one of π and $\overline{\pi}$ divides $n+i$. Therefore, either π or $\overline{\pi}$ is a greatest common divisor of p and $n+i$ in $\mathbb{Z}[i]$.
  - Thus, 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 $n+i$ 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.
  - Using modular exponentiation, we can verify that $3^{(p-1)/4} \equiv 1729 \pmod{p}$. Thus, our discussion above tells us that 1729 is a square root of -1 modulo p, and 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 $4(n_1+1)(n_2+1)\cdots(n_k+1)$.

[1] It may initially seem that computing $u^{(p-1)/4}$ modulo p would take a long time, but this calculation can be done rapidly using "successive squaring". As an example, to compute 2^{517} modulo 4457, we would compute 2^1, 2^2, 2^4, 2^8, ... , 2^{512} (each term is the square of the previous one, so these values are not hard to compute) and then evaluate $2^{517} = 2^{512} \cdot 2^4 \cdot 2^1$ modulo 4457.

- 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 $n = (1+i)^{2k} (\pi_1 \overline{\pi_1})^{n_1} \cdots (\pi_k \overline{\pi_k})^{n_k} q_1^{m_1} \cdots q_d^{m_d}$.
- 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 we gave above or by inspection), so the sixteen 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.

- As another application of our results, we can prove a classical characterization of the "Pythagorean triples" (triples of integers that represent the side lengths of a right triangle).
  - If $a^2 + b^2 = c^2$ for integers a, b, c, note that if two of a, b, c are divisible by a prime p, then so is the third. We can then "reduce" the triple (a, b, c) by dividing each term by p to obtain a new triple (a', b', c') with $(a')^2 + (b')^2 = (c')^2$.
  - For this reason it is sufficient to characterize the "primitive" Pythagorean triples with $\gcd(a, b, c) = 1$. For such triples, since a and b cannot both be odd (since then $a^2 + b^2 \equiv 2 \pmod 4$ cannot be a perfect square) we see that exactly one of a, b is even.

- Theorem (Pythagorean Triples): Every triple of positive integers (a, b, c) with $a^2 + b^2 = c^2$ with $\gcd(a, b, c) = 1$ and a even is of the form $(a, b, c) = (2st,\ s^2 - t^2,\ s^2 + t^2)$, for some relatively prime integers $s > t$ of opposite parity, and (conversely) any such triple is Pythagorean and primitive.
  - Proof: It is easy to see that $(2st)^2 + (s^2 - t^2)^2 = (s^2 + t^2)^2$ simply by multiplying out, and it is likewise not difficult to see that if s and t are relatively prime and have opposite parity, then $\gcd(s^2 - t^2, s^2 + t^2) = 1$ so this triple is primitive.
  - To show that (a, b, c) must be of the desired form, suppose that $a^2 + b^2 = c^2$, and factor the equation in $\mathbb{Z}[i]$ as $(a + bi)(a - bi) = c^2$.
  - We claim that $a + bi$ and $a - bi$ are relatively prime in $\mathbb{Z}[i]$: any gcd must divide $2x$ and $2y$, hence divide 2. However, $a + bi$ is not divisible by the prime $1 + i$, since a and b are of opposite parity.
  - Hence, since $a + bi$ and $a - bi$ are relatively prime and have product equal to a square, by the uniqueness of prime factorization in $\mathbb{Z}[i]$, there exists some $s + it \in \mathbb{Z}[i]$ and some unit $u \in \{1, i, -1, -i\}$ such that $a + bi = u(s + ti)^2$.
  - Multiplying out yields $a + bi = u\left[(s^2 - t^2) + (2st)i\right]$. Since a is even, b is odd, and both are positive, we must have $u = -i$ and $s > t$: then we see $a = 2st$, $b = s^2 - t^2$, and $c = s^2 + t^2$ as claimed.

4.4 Factorization of Ideals In Quadratic Integer Rings[2]

- 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.
  - 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.

[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`.

- 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[3]: 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 $\mathcal{O}_D$.

4.4.1 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.

[3] Recall that if I and J are ideals, then the product ideal IJ is defined to be the set of all finite products of an element of I with an element of J. In a commutative ring with 1, multiplication of ideals is commutative and associative, and if $I = (a_1, \ldots, a_n)$ and $J = (b_1, \ldots, b_m)$, then $IJ = (a_1b_1, \ldots, a_nb_m)$. To see this, observe that the product ideal IJ certainly contains all of these pairwise products, and conversely by distributing we see that any product of an element of I with an element of I lies in $(a_1b_1, \ldots, a_nb_m)$, hence so do sums of such products.

- 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.

4.4.2 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.

4.4.3 Applications of Unique Factorization 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 $\mathcal{O}_D$.

- 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 $(7) = (7, \sqrt{7})^2 = (\sqrt{7})^2$.

- As a final concluding remark, 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 finite-dimensional 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.

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