Contents

4 Unique Factorization and Applications

In this chapter, we extend the notion of a division algorithm to more general rings and then formalize the idea of when a ring possesses unique factorization. Our ultimate goal is to extend number-theoretic properties of $\mathbb{Z}$ to other number systems, so we then generalize the notion of modular arithmetic and establish the analogues of the Chinese remainder theorem, Fermat's little theorem, and Euler's theorem in the general setting.

We then apply our results in two rings relevant to number theory: the polynomial ring $F[x]$ and the Gaussian integer ring $\mathbb{Z}[i]$. In particular, we study the structure of the polynomials with coefficients in $\mathbb{Z}/p\mathbb{Z}$ and use the results to study finite fields and to characterize those m for which a primitive root exists modulo m. We also make in-depth study of the structure of the Gaussian integers, including giving a description of the modular arithmetic and prime factorization in $\mathbb{Z}[i]$. Along the way, we will also study finite fields from several perspectives, and establish Fermat's characterization of the integers that can be written as the sum of two squares.

4.1 Integral Domains, Euclidean Domains, and Unique Factorization

- Our goal in this section is to describe the class of rings that possess a division algorithm similar to that in $\mathbb{Z}$, and then establish that such rings have unique factorization.

4.1.1 Integral Domains

- Recall that in a general commutative ring R, we say that $a|b$ if there exists $k \in R$ such that $b = ak$.
- (Reminder) Definition: If R is a commutative ring, we say that $x \in R$ is a zero divisor if $x \neq 0$ and there exists a nonzero $y \in R$ such that $xy = 0$. (Note in particular that 0 is *not* a zero divisor!)
    - We originally defined zero divisors when discussing the ring structure of $\mathbb{Z}/m\mathbb{Z}$.
    - In $\mathbb{Z}/6\mathbb{Z}$, since $\overline{2} \cdot \overline{3} = \overline{4} \cdot \overline{3} = \overline{0}$, the residue classes represented by 2, 3, and 4 are zero divisors.
    - As a general philosophy, zero divisors can be a bit troublesome (at least, to a novice ring theorist), since they behave counter to one's natural intuition that products of nonzero elements should be nonzero.
- We recall a few important properties of zero divisors:
    - An integer a is a zero divisor modulo m if and only if $1 < \gcd(a, m) < m$. In particular, $\mathbb{Z}/m\mathbb{Z}$ contains zero divisors if and only if m is composite.
    - The ring $\mathbb{Z}/p\mathbb{Z}$ is a field (which, in particular, contains no zero divisors).
    - In a commutative ring with 1, a unit can never be a zero divisor.
- Definition: If R is a commutative ring with 1 that contains no zero divisors, R is called an integral domain (or often, just a domain).
    - Example: Any field is an integral domain. More generally, any ring that is a subset of a field is an integral domain: hence, the integers $\mathbb{Z}$ and the ring $\mathbb{Z}[\sqrt{D}]$ for any D are integral domains (since they are all subsets of the field of complex numbers $\mathbb{C}$).
    - Example: The ring of polynomials $F[x]$ where F is a field is also an integral domain.
- Integral domains generally behave more nicely than arbitrary rings, because they obey more of the laws of arithmetic that are familiar from $\mathbb{Z}$:
- Proposition (Properties of Integral Domains): If R is an integral domain, the following hold in R:
    1. Multiplication in R has a cancellation law: specifically, if $a \neq 0$ and $ab = ac$, then $b = c$.
        - Proof: Suppose that $ab = ac$: then by rearranging we see that $a(b - c) = 0$.
        - Then since R is an integral domain, we must either have $a = 0$ or $b - c = 0$. Hence, if $a \neq 0$, we must have $b - c = 0$ and so $b = c$.
    2. If $a|b$ and $b|a$ and a, b are nonzero, then $a = bu$ for some unit u.
        - Proof: Since $a|b$, there is some u with $a = bu$. Since $b|a$, there is some w with $b = aw$.
        - Multiplying the two equations gives $ab = abuw$, so $ab(1 - uw) = 0$. Since a and b are nonzero and R is a domain, we can cancel to see that $1 - uw = 0$, so that u is a unit.
    3. For any $m \neq 0$, $a|b$ is equivalent to $(ma)|(mb)$.
        - Proof: Follows directly from the cancellation property (1).
- The situation in property (2) of the proposition above is important enough that we give it a name:
- Definition: If R is a commutative ring with 1 and $a' = ua$ for some unit u, we say that a and a' are associates.
    - Notice that if a and a' are associates, then $a|a'$ and $a'|a$. For this reason, associates have very similar divisibility properties to one another.
    - Example: In $\mathbb{Z}$, the elements 2 and -2 are associates. Indeed, n and $-n$ are associates for any $n \in \mathbb{Z}$.
    - Example: In $\mathbb{Z}[i]$, the elements $1 + 2i$ and $2 - i$ are associates, because $2 - i = -i(1 + 2i)$.
    - Example: In $\mathbb{F}_3[x]$, the elements $x^2 + 2$ and $2x^2 + 1$ are associates, because $2x^2 + 1 = 2(x^2 + 2)$.
    - We will remark that "being associate" is an equivalence relation on R.

- Definition: Let $a, b \in R$ where R is an integral domain. We say d is a common divisor if $d|a$ and $d|b$, and 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 $d'|d$.
    - Example: 2 is a greatest common divisor of 14 and 20 in $\mathbb{Z}$, and -2 is also a greatest common divisor of 14 and 20. In particular, note that the greatest common divisor is no longer unique.
    - Example: In the polynomial ring $\mathbb{C}[x]$, a greatest common divisor of $x^2 - 4$ and $x^2 - 5x + 6$ is $x - 2$: we see $x - 2$ divides both $x^2 - 4 = (x - 2)(x + 2)$ and $x^2 - 5x + 6 = (x - 2)(x - 3)$, and there cannot be any common divisor of greater degree.
    - Note that we have not given a complete proof that these are actually greatest common divisors, since we would need to find all other common divisors and verify that they do divide the claimed gcd. (We will establish the correctness of these calculations shortly.)
- As an important warning, we will observe that in arbitrary integral domains, there can exist pairs (a, b) that do not possess a greatest common divisor.
- Example: Show that $2(1 + \sqrt{-5})$ and 6 do not possess a greatest common divisor in the ring $\mathbb{Z}[\sqrt{-5}]$.
    - First observe that 2 and $1 + \sqrt{-5}$ are both common divisors of $2(1 + \sqrt{-5})$ and 6.
    - Now we show using norms there is no element d that divides $2(1 + \sqrt{-5})$ and 6 that is also itself divisible by 2 and $1 + \sqrt{-5}$.
    - So suppose d does divide $2(1 + \sqrt{-5})$ and 6. Then necessarily $N(d)$ would divide $N(2 + 2\sqrt{-5}) = 24$ and $N(6) = 36$, so $N(d)$ 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, since there are no integer solutions to $a^2 + 5b^2 = 12$. Thus, there cannot be any such element d, meaning that $2(1 + \sqrt{-5})$ and 6 do not possess a greatest common divisor.
- If two elements do have a greatest common divisor, then it is unique up to taking associates:
- Proposition (GCDs and Associates): Let R be a commutative ring with 1 and $a, b \in R$. If d_1 and d_2 are both greatest common divisors of a and b, then d_1 and d_2 are associates. Conversely, if d is a greatest common divisor of a and b, then so is any associate of d.
    - Proof: Since d_1 is a gcd, and d_2 is a common divisor, we see $d_1|d_2$, and similarly $d_2|d_1$. By property (2) of divisibility in integral domains above, this implies $d_1 = ud_2$ for a unit u, so d_1 and d_2 are associates.
    - For the other statement, suppose that d is a greatest common divisor of a and b and ud is any associate of d (where by assumption u is a unit).
    - Then since $d|a$, there exists $c \in R$ with $a = dc$, and so $a = (cu^{-1})(ud)$. This means $(ud)|a$. Likewise, $(ud)|b$, so ud is also a common divisor of a and b.
    - Also, if d' is any other common divisor, then $d'|d$ by assumption. Since $d|(ud)$ this means $d'|(ud)$, so every common divisor divides ud. Hence ud is also a greatest common divisor of a and b.
- In particular we can also define the analogue of relatively prime elements in an arbitrary domain:
- Definition: If R is a commutative ring with 1 and 1 is a greatest common divisor of r and s, we say r and s are relatively prime.
    - By the proposition above, r and s are relatively prime if and only if they have a greatest common divisor that is a unit.
    - Example: 2 and 5 are relatively prime in $\mathbb{Z}$ since they have a gcd of 1.

4.1.2 Euclidean Domains and Division Algorithms

- We now discuss what it means for an integral domain to possess a "division algorithm".
- Definition: If R is a domain, any function $N : R \to \mathbb{N} \cup \{0\}$ 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.
  - We will mention that the norm we have defined on the rings $\mathbb{Z}[\sqrt{D}]$ is not technically a norm under this definition (it is, however, if we take the absolute value). We will leave the exact choice of whether the absolute value is included up to context.
- 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.
  - 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, because if we set $N(n) = |n|$, then, as we have already proven, the standard division algorithm allows us to write $a = qb + r$ with either $r = 0$ or $|r| < |b|$.
- 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 R):
- 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).
- In precisely the same as in $\mathbb{Z}$, we can use the Euclidean algorithm to establish the existence of greatest common divisors and give a procedure for calculating them:
- 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. Furthermore, there exist elements $x, y \in R$ such that $d = ax + by$.
  - The ideas in the proof are the same as for the proof over $\mathbb{Z}$.
  - 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 $x_i, y_i \in R$.

- Corollary: Any two elements in a Euclidean domain always possess a greatest common divisor.
- With the motivation for our choice of definition in hand, we can now give our two fundamental examples of Euclidean domains. First, we describe the division algorithm in $\mathbb{Z}[i]$:
- 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[1] $q, r \in \mathbb{Z}[i]$ 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, we see that $\left|\frac{r}{c+di}\right|^2 = |(x-s)+(y-t)i|^2 = (x-s)^2+(y-t)^2 \leq \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$.
  - Clearing the denominator yields $N(r) = |r|^2 \leq \frac{1}{2}|c+di|^2 = \frac{1}{2}N(c+di)$, as desired.
- By using the Euclidean algorithm (by computing the quotient and remainder as described in the proof above) we may now compute greatest common divisors in $\mathbb{Z}[i]$ and write them as explicit linear combinations, just as we did in $\mathbb{Z}$:
- Example: Find a greatest common divisor of $50-50i$ and $43-i$ in $\mathbb{Z}[i]$, and write it as an explicit linear combination of $50-50i$ and $43-i$.
  - We use the Euclidean algorithm. Dividing $43-i$ into $50-50i$ yields $\frac{50-50i}{43-i} = \frac{44}{37} - \frac{42}{37}i$, 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 $\frac{43-i}{8-6i} = \frac{7}{2}+\frac{5}{2}i$, 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 $\frac{8-6i}{7+i} = 1-i$, 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 $7+i = \boxed{(-3-2i)\cdot(50-50i) + (6-i)\cdot(43-i)}$.

[1] For the rings $\mathbb{Z}[\sqrt{D}]$ in general, the function $N(a+b\sqrt{D}) = |a^2-Db^2|$ is a norm, but it does not in general give a division algorithm. Only for certain small values of $|D|$, like $D=-1$, will this function allow us to construct quotients and remainders where the remainder is smaller (in norm) than the element being divided by.

- Example: Find a greatest common divisor of $11+18i$ and $8-3i$ in $\mathbb{Z}[i]$, and write it as an explicit linear combination of $11+18i$ and $8-3i$.
  - We use the Euclidean algorithm:
$$\begin{aligned} 11+18i &= 2i\cdot(8-3i)+(5+2i) \\ 8-3i &= (1-i)\cdot(5+2i)+1 \\ 5+2i &= (5+2i)\cdot 1 \end{aligned}$$
  - The last nonzero remainder is $\boxed{1}$ so it is a gcd. To express the gcd as a linear combination, we solve for the remainders:
$$\begin{aligned} 5+2i &= (11+18i)-2i\cdot(8-3i) \\ 1 &= (8-3i)-(1-i)\cdot(5+2i) \\ &= (8-3i)-(1-i)(11+18i)+2i(1-i)(8-3i) \\ &= (-1+i)(11+18i)+(3+2i)(8-3i) \end{aligned}$$
  and so we have $1=\boxed{(-1+i)(11+18i)+(3+2i)(8-3i)}$.
- We now show that $F[x]$ is Euclidean:
- Proposition ($F[x]$ is Euclidean): If F is any field, the ring of polynomials $F[x]$ in the variable x with coefficients in F is a Euclidean domain, under the norm given by $N(p(x))=\deg(p)$.
  - The idea is simply to show the validity of polynomial long division. The reason we require F to be a field is that we need to be able to divide by arbitrary nonzero coefficients to be able to perform the divisions. (Over $\mathbb{Z}$, for instance, we cannot divide x^2 by $2x$ and get a remainder that is a constant polynomial.)
  - Explicitly, we will show that if $a(x)$ and $b(x)$ are polynomials with $b(x)\neq 0$, then there exist $q(x)$ and $r(x)$ such that $a(x)=q(x)b(x)+r(x)$, and either $r(x)=0$ or $\deg(r)<\deg(b)$.
  - Proof: We prove this by induction on the degree n of $a(x)$. The base case is trivial, as we may take $q=r=0$ if $a=0$.
  - Now suppose the result holds for all polynomials $a(x)$ of degree $\leq n-1$. If $\deg(b)>\deg(a)$ then we can simply take $q=0$ and $r=a$, so now also assume $\deg(b)\leq\deg(a)$.
  - Write $a(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_0$ and $b(x)=b_mx^m+\cdots+b_0$, where $b_m\neq 0$ since $b(x)\neq 0$.
  - Observe that the polynomial $a^\dagger(x)=a(x)-\dfrac{a_n}{b_m}x^{n-m}b(x)$ has degree less than n, since we have cancelled the leading term of $a(x)$. (Here we are using the fact that F is a field, so that $\dfrac{a_n}{b_m}$ also lies in F.)
  - By the induction hypothesis, $a^\dagger(x)=q^\dagger(x)b(x)+r^\dagger(x)$ for some $q^\dagger(x)$ and $r^\dagger(x)$ with $r^\dagger=0$ or $\deg(r^\dagger)<\deg(b)$.
  - Then $a(x)=\left[q^\dagger(x)+\dfrac{a_n}{b_m}x^{n-m}\right]b(x)+r^\dagger(x)$, so $q(x)=q^\dagger(x)+\dfrac{a_n}{b_m}x^{n-m}$ and $r(x)=r^\dagger(x)$ satisfy all of the requirements.
  - Remark: It is also straightforward to see that the quotient and remainder are unique under the requirement that $\deg(r)<\deg(b)$, by observing that if $a=qb+r=q'b+r'$, then $r-r'$ has degree less than $\deg(b)$ but is also divisible by $b(x)$, hence must be zero.
- Example: Find a greatest common divisor $d(x)$ of the polynomials $p=x^6+2$ and $q=x^8+2$ in $\mathbb{F}_3[x]$, and then write the gcd as a linear combination of p and q.
  - We apply the Euclidean algorithm: we have
$$\begin{aligned} x^8+2 &= x^2(x^6+2)+(x^2+2) \\ x^6+2 &= (x^4+x^2+1)(x^2+2) \end{aligned}$$
  and so the last nonzero remainder is $\boxed{x^2+2}$.

- By back-solving, we see that $x^2+2 = \boxed{1\cdot(x^8+2) - x^2(x^6+2)}$.

- When performing the Euclidean algorithm in $F[x]$, the coefficients can often become quite large or complicated:

- Example: Find a greatest common divisor $d(x)$ of the polynomials $p = x^3+7x^2+9x-2$ and $q = x^2+4x$ in $\mathbb{R}[x]$, and then write the gcd as a linear combination of p and q.

  - We apply the Euclidean algorithm: we have

$$\begin{aligned} x^3+7x^2+9x-2 &= (x+3)(x^2+4x) + (-3x-2) \\ x^2+4x &= (-\frac{10}{9} - \frac{1}{3}x)(-3x-2) + (-20/9) \\ -3x-2 &= \frac{27x+6}{20}(-20/9) \end{aligned}$$

  and so the last nonzero remainder is $-20/9$. Thus, by rescaling, we see that the gcd is $\boxed{1}$.

  - By back-solving, we see that

$$\begin{aligned} -3x-2 &= 1\cdot(x^3+7x^2+9x-2) - (x+3)\cdot(x^2+4x) \\ -20/9 &= x^2+4x + (\frac{10}{9}+\frac{1}{3}x)(-3x-2) \\ &= (\frac{10}{9}+\frac{1}{3}x)\cdot(x^3+7x^2+9x-2) - (\frac{7}{3}+\frac{19}{9}x+\frac{1}{3}x^2)\cdot(x^2+4x) \end{aligned}$$

  and thus by rescaling, we obtain $1 = \boxed{(-\frac{1}{2}-\frac{3}{20})\cdot(x^3+7x^2+9x-2) + (\frac{21}{20}+\frac{19}{20}x+\frac{3}{20}x^2)\cdot(x^2+4x)}$.

4.1.3 Irreducible and Prime Elements

- Now that we have given the definition of a general ring having a "division algorithm", we would like to discuss when a ring has unique factorization.

  - In order to do this, we first need a notion generalizing the idea of a prime number in $\mathbb{Z}$: namely, an element that does not have any nontrivial divisors.

- Definition: If R is an integral domain, a nonzero element $a \in R$ is irreducible if it is not a unit and, for any "factorization" $a = bc$ with $b, c \in R$, one of b and c must be a unit.

  - Example: The irreducible elements of $\mathbb{Z}$ are precisely the prime numbers (and their negatives).
  - 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 $\mathbb{Z}[i]$.
  - In $\mathbb{Z}[\sqrt{D}]$, we can often test for reducibility using norms.
  - Example: The element $2+i$ is irreducible in $\mathbb{Z}[i]$: 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.
  - Important Warning: Whether a given element is irreducible depends on the ring R of which it is an element. For example, 5 is irreducible in the ring $\mathbb{Z}$, but 5 is not irreducible in the ring $\mathbb{Z}[i]$ because in this ring we can write $5 = (2+i)(2-i)$ and neither of these elements is a unit.

- The irreducible elements of $F[x]$ are the irreducible polynomials of positive degree: namely, the polynomials that cannot be factored into a product of polynomials of smaller positive degree.

- In polynomial rings, we can often use degrees to see immediately that a given polynomial is irreducible, since if $f = gh$ is a nontrivial factorization, then $\deg(f) = \deg(g) + \deg(h)$, where $\deg(g)$ and $\deg(h)$ must both be positive.
- Example: Any polynomial of degree 1 is irreducible, since if $f = gh$ with $\deg(g), \deg(h)$ positive, then $\deg(f) = \deg(g) + \deg(h) \geq 2$.
- Example: The polynomial $x^2 + x + 1$ is irreducible in $\mathbb{F}_2[x]$, since the only possible factorizations would be $x \cdot x$, $x \cdot (x+1)$, or $(x+1) \cdot (x+1)$, and none of these is equal to $x^2 + x + 1$.
- Example: The polynomial x^4+4 is reducible in $\mathbb{Q}[x]$, since we can write $x^4+4 = (x^2+2x+2)(x^2-2x+2)$.
- Example: The polynomial $x^2 + 1$ is irreducible in $\mathbb{R}[x]$, since there is no way to write it as the product of two linear polynomials with real coefficients.
- Important Warning: Whether a given polynomial is irreducible depends on the ring $F[x]$ of which it is an element. For example, x^2+1 is irreducible in $\mathbb{R}[x]$ but not in $\mathbb{C}[x]$, since we can write $x^2+1 = (x+i)(x-i)$ in $\mathbb{C}[x]$.

- An irreducible element behaves much like a prime number in $\mathbb{Z}$. However, there is a separate notion of a prime element in a general domain:
- Definition: If R is an integral domain, a nonzero element $p \in R$ is prime if it is not a unit and, for any $a, b \in R$ such that $p|ab$, it must be the case that $p|a$ or $p|b$.
  - 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.
  - Based on these two examples, it may seem that irreducible and prime elements are always the same. They are indeed closely related, but they do not always coincide:
  - Non-Example: Although the element 2 is irreducible in $\mathbb{Z}[\sqrt{-5}]$, it is not prime: 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.
- In fact, prime elements are always irreducible:
- Proposition (Primes are Irreducible): If R is an integral domain and $p \in R$ is a prime element, then p is irreducible.
  - Proof: Suppose p is prime and has a factorization $p = bc$.
  - By definition, it must be the case that $p|b$ or $p|c$; by relabeling assume $p|b$, with $b = pu$ for some u.
  - Then $p = puc$ so $p(1 - uc) = 0$. Cancelling p yields $uc = 1$, so c is a unit.
  - Thus, in any factorization $p = bc$, at least one term must be a unit, and this means p is irreducible.

4.1.4 Unique Factorization Domains

- We would like to say a ring possesses unique factorization if every nonzero element can be written uniquely as the product of irreducibles. However, there are some issues with this attempted definition.
  - To illustrate, observe that in the Gaussian integers, we can write $5 = (2 - i)(2 + i) = (1 + 2i)(1 - 2i)$.
  - It would seem that these are two different factorizations, but we should really consider them the same, because all we have done is moved some units around: $(2 - i) \cdot i = 1 + 2i$ and $(2 + i) \cdot (-i) = 1 - 2i$.
  - We should declare that two factorizations are equivalent if the only differences between them are by moving units around, which is equivalent to replacing elements with associates.
- Definition: An integral domain R is a unique factorization domain (UFD) if (i) every nonzero nonunit $r \in R$ can be written as a finite product of irreducibles $r = p_1 p_2 \cdots p_k$, and (ii) such a factorization is unique up to associates: if $r = q_1 q_2 \cdots q_d$ is some other factorization, then $d = k$ and there is some reordering of the factors such that p_i is associate to q_i for each $1 \leq i \leq k$.
- Our main result is the following:

- Theorem (Euclidean Domains are UFDs): If R is a Euclidean domain, then R is a unique factorization domain.
  - The main ideas of the proof are the same as those over $\mathbb{Z}$, but they are a bit obfuscated by some of the technical difficulties. The existence portion of the proof contains precisely the same ideas as in our characterization of the gcd of a collection of integers as the minimal positive linear combination of those integers. The uniqueness portion is likewise essentially the same as for $\mathbb{Z}$, namely, an induction argument on the number of irreducible terms in a factorization.
  - Proof (Existence): Let R be a Euclidean domain and r a nonzero nonunit.
    - If r is irreducible, we are done. Otherwise, by definition 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 claim that this process must terminate eventually: otherwise, there necessarily exists 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.
    - Consider the set I of all (finite) R-linear combinations $I = \{r_1 x_1 + r_2 x_2 + \cdots + r_k x_k : k \geq 1, r_i \in R\}$, and let $y \in I$ be a nonzero element in I of minimal norm.
    - We claim that every element in I is divisible by y: otherwise, if there were some element $s \in I$ with y not dividing s, applying the division algorithm to write $s = qy + r^\dagger$ would yield the element $r^\dagger = (s - qy) \in I$ of smaller norm than y, contradiction.
    - But now since $y \in I$, we can write $y = r_1 x_1 + \cdots + r_d x_d$ for some d. Since $x_d | x_{d-1} | \cdots | x_1$, the LHS is a multiple of x_d, meaning $x_d | y$.
    - But now since $x_d | y$ and $y | x_{d+1}$ we conclude $x_d | x_{d+1}$. But by assumption, $x_{d+1} | x_d$, meaning that they are associates; this is a contradiction.
    - Hence the factoring process must terminate, as claimed.
  - Proof (Uniqueness): Let R be a Euclidean domain and r be a nonzero nonunit. We prove the factorization of r is unique by induction on the number of irreducible factors of $r = p_1 p_2 \cdots p_d$.
    - If $d = 0$, then r is a unit. If r had some other factorization $r = qc$ with q irreducible, then q would divide a unit, hence be a unit (impossible).
    - Now suppose $d \geq 1$ 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, 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.
    - Then $q_i = p_1 u$ for some u: then since q_i is irreducible (and p_1 is not a unit), u must be a unit, so p_1 and q_i are associates.
    - Cancelling then yields the equation $p_2 \cdots p_d = (u q_2) \cdots q_k$, 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.
  - Remark (for those who like ring theory): The set I from the existence proof is an ideal. The underlying idea of the proof given above is to show that any ideal in a Euclidean domain is actually principal (generated by a single element): in other words, that Euclidean domains are "principal ideal domains". In fact, it is actually true that a Euclidean domain is a principal ideal domain, and that any principal ideal domain is a unique factorization domain.

- Corollary: The Gaussian integers $\mathbb{Z}[i]$ are a unique factorization domain, as is the polynomial ring $F[x]$ for any field F.

- We will note that, although most of the unique factorization domains we will discuss are also Euclidean domains, there exist UFDs that are not Euclidean domains[2].

- We often prefer to speak of a "prime factorization", rather than a "factorization into irreducibles". In a Euclidean domain, these are equivalent:

[2]One example is the ring $\mathbb{C}[x, y]$ of polynomials in the two variables x and y, with coefficients in $\mathbb{C}$. There is no Euclidean division algorithm in this ring (the degree map is not a Euclidean norm since, e.g., there is no way to divide y^2 by x and obtain a remainder of degree zero). However, polynomials in $\mathbb{C}[x, y]$ can still be factored uniquely into a product of irreducible terms.

- Proposition (Primes and Irreducibles in Euclidean Domains): If R is a Euclidean domain, then $p \in R$ is prime if and only if it is irreducible.
    - Proof: We showed that primes are irreducible earlier. Now suppose p is irreducible and that $p|ab$: we wish to show that $p|a$ or $p|b$.
    - If $p|a$, we are done, so suppose $p \nmid a$, and let d be a gcd of p and a, which exists since R is a Euclidean domain.
    - By hypothesis, d divides p, so (since p is irreducible) either d is a unit, or $d = up$ for some unit u: however, the latter cannot happen, because then up (hence p) would divide a. Hence d is a unit, say with inverse e.
    - By the Euclidean algorithm, we see that there exist x and y such that $xp + ya = d$. Multiplying by be and regrouping the terms yields $(bce)p + ey(ab) = (de)b = b$. Since p divides both terms on the left-hand side, we conclude $p|b$.

4.2 Modular Arithmetic in Euclidean Domains

- We have previously described the division algorithm over $\mathbb{Z}$ and used it to study modular arithmetic in $\mathbb{Z}$. The goal of this section is to show that there is a meaningful extension of the notion of "modular arithmetic" modulo a general element r in a Euclidean domain R, and then to establish the analogues of the major results from in $\mathbb{Z}$: the Chinese remainder theorem, and the theorems of Fermat and Euler.
    - Our primary interest is when R is $\mathbb{Z}[i]$ or $F[x]$, for F a field.
    - However, many of the notions will hold in general, and so we will work in the general setting whenever possible. We will see that almost all of the proofs are exactly the same as over $\mathbb{Z}$.
    - We will also, when possible, remark when the results we prove hold (or fail to hold) for more general classes of rings.
    - Remark (for those who like ring theory): All of what we do here is subsumed by the theory of ideals in a general ring R, and our construction of "modular arithmetic" is a special case of the quotient of a ring by an ideal. (Specifically, we are studying the quotient rings of the form R/I where I is a principal ideal. Every ideal is principal in a Euclidean domain, so we do not lose anything here by studying quotients without speaking of ideals explicitly.)

4.2.1 Modular Congruences and Residue Classes

- Our underlying definition of modular congruences and residue classes are exactly the same as over $\mathbb{Z}$:
- Definition: Let R be a commutative ring with 1. If $a, b, r \in R$, we say that a is congruent to b modulo c, written $a \equiv b \pmod{c}$, if $c|(b-a)$. The residue class of a modulo r, denoted $\overline{a}$, is the set $S = \{a + dr : d \in R\}$ of all elements in R congruent to a modulo r.
    - Example: In $\mathbb{Z}[i]$, it is true that $13-3i \equiv 2-i$ modulo $3+4i$, because $(13-3i)-(2-i) = (1-2i)(3+4i)$.
    - Example: In $\mathbb{F}_2[x]$, it is true that $x^3 + x \equiv x + 1$ modulo $x^2 + x + 1$, because $(x^3 + x) - (x + 1) = (x + 1) \cdot (x^2 + x + 1)$.
- All of the properties of residue classes and congruences from $\mathbb{Z}$ extend to R:
- Proposition (Congruences and Residue Classes): Let R be an integral domain. For any $r, a, b, c, d \in R$, the following are true:
    1. We have $a \equiv a \pmod{r}$, $a \equiv b \pmod{r}$ if and only if $b \equiv a \pmod{r}$, and if $a \equiv b \pmod{r}$ and $b \equiv c \pmod{r}$ then $a \equiv c \pmod{r}$.
    2. If $a \equiv b \pmod{r}$ and $c \equiv d \pmod{r}$, then $a + c \equiv b + d \pmod{r}$ and $ac \equiv bd \pmod{r}$.
    3. We have $a \equiv b \pmod{r}$ if and only if $\overline{a} = \overline{b}$.
    4. Two residue classes modulo r are either disjoint or identical.

- Proof: The proofs of all of these statements are the same as over $\mathbb{Z}$.

- Proposition (Modular Arithmetic): The set R/rR consisting of all residue classes in R modulo r forms a ring under the addition and multiplication operations $\overline{a} + \overline{b} = \overline{a+b}$ and $\overline{a} \cdot \overline{b} = \overline{a \cdot b}$.
  - Proof: The most difficult part is showing that the addition and multiplication operations are well-defined: that if we choose different elements $a' \in \overline{a}$ and $b' \in \overline{b}$, the residue class of $a' + b'$ is the same as that of $a + b$, and similarly for the product.
  - Explicitly, suppose $a' \in \overline{a}$ and $b' \in \overline{b}$. Then there exists $k_1 \in R$ such that $a' = a + k_1 r$ and also $k_2 \in R$ such that $b' = b + k_2 r$.
  - Then $a' + b' = (a + b) + r(k_1 + k_2)$, and since these differ by a multiple of r, we see that $\overline{a' + b'} = \overline{a + b}$, so addition is well-defined.
  - Similarly, $a'b' = (a + k_1 r)(b + k_2 r) = ab + r(k_1 b + k_2 a + k_1 k_2 r)$, so $\overline{a'b'} = \overline{ab}$, and multiplication is also well-defined.
  - Then the ring axioms [R1]-[R8] all follow directly from their counterparts in R. The additive identity in R/rR is $\overline{0}$, the additive inverse of $\overline{a}$ is $\overline{-a}$, and the multiplicative identity is $\overline{1}$.

- Over $\mathbb{Z}$, we usually work with a specific collection of representatives for the residue classes modulo m, generally the integers 0 through $m - 1$.
  - In a general ring, there is not usually a natural choice for residue class representatives.
  - Or, if there does happen to be a good choice, it is not always obvious what that choice is. (For example, try coming up with a natural choice of residue class representatives for $\mathbb{Z}[i]$ modulo $3 + 4i$.)
  - Unfortunately, this lack of an obvious choice makes it somewhat difficult to give concrete examples in situations that require a complete list of representatives. We will later describe ways to find a set of representatives for $\mathbb{Z}[i]$ modulo a prime p, and for $F[x]$ modulo an arbitrary polynomial.

4.2.2 Arithmetic in R/rR

- In the ring $F[x]$, where F is a field, we do get a natural collection of residue class representatives arising from the division algorithm.

- Proposition (Residue Classes for $F[x]$): If $R = F[x]$ and $q(x) \in R$ is a polynomial of degree d, then the polynomials in $F[x]$ of degree $\leq d - 1$ are a full set of residue class representatives for R/qR.
  - Proof: By the division algorithm, every polynomial is congruent modulo $q(x)$ to some polynomial of degree less than d, namely, to the remainder after dividing by $q(x)$.
  - Conversely, each of these residue classes is distinct, because two distinct polynomials in $F[x]$ of degree less than d cannot be congruent modulo $q(x)$: if they were, their difference would be a multiple of $q(x)$ of degree less than d, but the only such multiple is 0.

- Example: Describe the addition and multiplication in the ring R/qR, where $R = \mathbb{R}[x]$ and $q(x) = x^2 + 1$.
  - From the proposition above, since q has degree 2, the elements of R/qR are are of the form $a + b\overline{x}$ where $a, b \in \mathbb{R}$.
  - The addition is simply addition of polynomials: $(a + b\overline{x}) + (c + d\overline{x}) = (a + c) + (b + d)\overline{x}$.
  - The multiplication is also simply multiplication of polynomials, subject to the relation $\overline{x}^2 + 1 = 0$.
  - Thus, in general we can write $(a + b\overline{x}) \cdot (c + d\overline{x}) = ac + (bc + ad)\overline{x} + bd\overline{x}^2 = (ac - bd) + (bc + ad)\overline{x}$.
  - This multiplication should look very familiar: in fact, it is exactly the same as the multiplication of complex numbers $(a + bi) \cdot (c + di) = (ac - bd) + (bc + ad)i$.
  - There is an obvious reason for this: the ring $\mathbb{R}/q\mathbb{R}$ is really just the complex numbers $\mathbb{C}$, where instead of using $i^2 = -1$, we say $\overline{x}^2 = -1$. (In the language of algebra, we would say that R/qR and $\mathbb{C}$ are isomorphic as rings, meaning that their ring structures are exactly the same: we have just labeled the elements differently.)

- In particular, since $\mathbb{C}$ is a field, we see that R/qR is also a field.

- When F is an infinite field, there will be infinitely many residue classes in R/pR, so except in nice cases it is difficult to write out the multiplication explicitly. However, we can easily construct addition and multiplication tables when F is finite.

- Example: With $R = \mathbb{F}_2[x]$, here are the addition and multiplication tables for R/pR with $p = x^2$:

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$\cdot$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	0	x
$x+1$	0	$x+1$	x	1

  - Notice that this ring has a zero divisor (namely x), and that the elements 1 and $x+1$ are units. Notice that $p(x) = x^2$ is reducible in R, since it has the factorization $x^2 = x \cdot x$.

- Example: With $R = \mathbb{F}_2[x]$, here are the addition and multiplication tables for R/pR with $p = x^2 + x + 1$:

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$\cdot$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

  - Notice that this ring is a field, since every nonzero residue class is a unit. Observe also that the polynomial $p(x) = x^2 + x + 1$ is irreducible in R, since it has no roots.

- Example: With $R = \mathbb{Z}[i]$, here are the addition and multiplication tables for R/rR with $r = 2 + i$:

$+$	0	i	-1	$-i$	1
0	0	i	-1	$-i$	1
i	i	1	$-i$	0	-1
-1	-1	$-i$	i	1	0
$-i$	$-i$	0	1	-1	i
1	1	-1	0	i	$-i$

$\cdot$	0	i	-1	$-i$	1
0	0	0	0	0	0
i	0	-1	$-i$	1	i
-1	0	$-i$	1	i	-1
$-i$	0	1	i	1	$-i$
1	0	i	-1	$-i$	1

  - It is less obvious that these elements do give representatives of all of the residue classes: this follows because any possible remainder x upon dividing by $2+i$ must have $N(x) \leq \frac{1}{2}N(2+i) = \frac{5}{2}$, so the only possible remainders are the elements of norm 0, 1, and 2: namely, 0, ± 1, $\pm i$, and $\pm 1 \pm i$. The last four are all seen to be equivalent to the first five, since, for example $1+i \equiv -i \pmod{2+i}$ and $1-i \equiv i \pmod{2+i}$), and the first five all yield distinct residue classes since no two of them are congruent modulo $2+i$.
  - Thus, there are 5 residue classes modulo $2+i$: $\overline{0}$, $\overline{i}$, $\overline{-1}$, $\overline{-i}$, and $\overline{1}$.
  - Notice that this ring is a field, since every nonzero residue class is a unit. As we have also shown previously, $2+i$ is irreducible in R.

4.2.3 Units and Zero Divisors in R/rR

- As suggested by the examples above, and also by the analogies between $\mathbb{Z}/m\mathbb{Z}$ and R/rR, we can characterize the units and zero divisors in R/rR:

- Proposition (Units in R/rR): If R is a Euclidean domain, an element $s \in R$ is a unit in R/rR if and only r and s are relatively prime, and an element $s \in R$ is a zero divisor in R/rR whenever $\overline{s} \neq \overline{0}$ and r and s are not relatively prime.

  - Proof: If r and s are relatively prime, then since R is a Euclidean domain, there exist $a, b \in R$ such that $ar + bs = 1$. Then, modulo r, we have $\overline{b} \cdot \overline{s} = \overline{1}$, meaning that s is a unit in R/rR.
  - Conversely, suppose that s is a unit in R/rR: then there exists some b such that $\overline{b} \cdot \overline{s} = \overline{1}$ in R/rR.

- This means there exists some $a \in R$ with $bs = 1 - ar$, which is to say, with $ar + bs = 1$. Then since any gcd of r and s must divide $ar + bs$, we conclude that any gcd must be a unit. Since all gcds are associates, we conclude 1 is a gcd of r and s.
- For the second statement, if s is a zero divisor than it cannot be a unit, so by what we just showed, this means r and s cannot be relatively prime. Conversely, suppose that d is a greatest common divisor of r and s and d is not a unit: then $\overline{s} \cdot \overline{r/d} = \overline{s/d} \cdot \overline{r} = \overline{0}$ in R/rR, and since $\overline{r/d}$ is not zero modulo $\overline{r}$ since d is not a unit, this means $\overline{s}$ is a zero divisor.

- Just as in $\mathbb{Z}$, the proof of the result above gives us a procedure for computing the inverse of a unit u in R/rR (namely, by using the Euclidean algorithm to write 1 as a linear combination of u and r).
  - There is an additional minor wrinkle in that the result of the Euclidean algorithm may yield a gcd that is not 1 but rather some other unit in R: in such a case we need only scale both sides of the resulting linear combination by the inverse of that unit to obtain a linear combination of 1.

- Example: In $\mathbb{Z}[i]$, show that $7 - 2i$ is a unit modulo $11 + 8i$ and find its multiplicative inverse.
  - We apply the Euclidean algorithm:

$$\begin{aligned} 11 + 8i &= (1 + i)(7 - 2i) + (2 + 3i) \\ 7 - 2i &= (1 - 2i)(2 + 3i) + (-1 - i) \\ 2 + 3i &= -2(-1 - i) + i \\ -1 - i &= (-1 + i)(i) \end{aligned}$$

    The greatest common divisor is the last nonzero remainder of i. Since this is associate to 1, we see that $7 - 2i$ is a unit modulo $11 + 8i$.
  - To compute the inverse we solve for the remainders:

$$\begin{aligned} 2 + 3i &= 1(11 + 8i) + (-1 - i)(7 - 2i) \\ -1 - i &= 7 - 2i - (1 - 2i)(2 + 3i) \\ &= (-1 + 2i)(11 + 8i) + (4 - i)(7 - 2i) \\ i &= 2 + 3i + 2(-1 - i) \\ &= (-1 + 4i)(11 + 8i) + (7 - 3i)(7 - 2i) \end{aligned}$$

    and so $i = (-1+4i)(11+8i)+(7-3i)(7-2i)$. Multiplying by $-i$ yields $1 = (4+i)(11+8i)+(-3-7i)(7-2i)$, and then reducing modulo $11 + 8i$ yields $(-3 - 7i) \cdot (7 - 2i) \equiv 1 \pmod{11 + 8i}$.
  - Hence the inverse of $7 - 2i$ modulo $11 + 8i$ is $\boxed{-3 - 7i}$.

- Example: In $\mathbb{Z}[i]$, show that $7 - 2i$ is a unit modulo $11 + 8i$ and find its multiplicative inverse.
  - We apply the Euclidean algorithm:

$$\begin{aligned} 11 + 8i &= (1 + i)(7 - 2i) + (2 + 3i) \\ 7 - 2i &= (1 - 2i)(2 + 3i) + (-1 - i) \\ 2 + 3i &= -2(-1 - i) + i \\ -1 - i &= (-1 + i)(i) \end{aligned}$$

    The greatest common divisor is the last nonzero remainder of i. Since this is associate to 1, we see that $7 - 2i$ is a unit modulo $11 + 8i$.
  - To compute the inverse we solve for the remainders:

$$\begin{aligned} 2 + 3i &= 1(11 + 8i) + (-1 - i)(7 - 2i) \\ -1 - i &= 7 - 2i - (1 - 2i)(2 + 3i) \\ &= (-1 + 2i)(11 + 8i) + (4 - i)(7 - 2i) \\ i &= 2 + 3i + 2(-1 - i) \\ &= (-1 + 4i)(11 + 8i) + (7 - 3i)(7 - 2i) \end{aligned}$$

and so $i = (-1+4i)(11+8i)+(7-3i)(7-2i)$. Multiplying by $-i$ yields $1 = (4+i)(11+8i)+(-3-7i)(7-2i)$, and then reducing modulo $11+8i$ yields $(-3-7i)\cdot(7-2i) \equiv 1 \pmod{11+8i}$.

- Hence the inverse of $7-2i$ modulo $11+8i$ is $\boxed{-3-7i}$.

- Example: For $R = \mathbb{F}_5[x]$, find the multiplicative inverse of x^2+2 modulo x^3+1.

  - First we apply the Euclidean algorithm in R:

$$\begin{aligned} x^3+1 &= x\cdot(x^2+2)+(3x+1) \\ x^2+2 &= (2x+1)\cdot(3x+1)+1 \\ 3x+1 &= (3x+1)\cdot 1 \end{aligned}$$

    and so the gcd of x^2+2 and x^3+1 is 1. Hence x^2+2 is indeed a unit modulo x^3+1.

  - To compute the inverse we solve for the remainders:

$$\begin{aligned} 3x+1 &= (x^3+1)-x\cdot(x^2+2) \\ 1 &= (x^2+2)-(2x+1)(3x+1) = (2x^2+x+1)(x^2+2)-(2x+1)(x^3+1) \end{aligned}$$

    and thus by reducing both sides modulo x^3+1, we see that the multiplicative inverse of x^2+2 modulo x^3+1 is $\boxed{2x^2+x+1}$.

- One of the other nice properties of $\mathbb{Z}/m\mathbb{Z}$ is that if p is prime, then $\mathbb{Z}/p\mathbb{Z}$ is actually a field. This remains true if we replace $\mathbb{Z}$ with an arbitrary Euclidean domain:

- Proposition (R/pR and Fields): If R is a Euclidean domain, the element $p \in R$ is a prime element (equivalently, irreducible) if and only if R/pR is a field.

  - Proof: Suppose p is a prime element. If $p|a$, then $a \equiv 0 \pmod{p}$, so $\overline{a} = \overline{0}$. Now suppose that p does not divide a.
  - Then because p is prime (hence irreducible), the only possible common divisors of a and p are units. This means a and p are relatively prime, so a is a unit modulo p. Thus, every nonzero element in R/pR is a unit, so R/pR is a field.
  - Conversely, suppose R/pR is a field. If $a, b \in R$ are such that $p|ab$, then $ab \equiv 0 \pmod{p}$.
  - Since R/pR is a field, it has no zero divisors, meaning that $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$, which is to say, $p|a$ or $p|b$. Thus, p is a prime element of R.
  - Remark: This proposition is *not* true if R is only assumed to be a general commutative ring with 1, or even a unique factorization domain. (The unique factorization domain $R = \mathbb{C}[x,y]$ is a counterexample: the element x is prime, but the ring R/xR is not a field, since $\overline{y}$ has no multiplicative inverse there.) The correct equivalence in that case is "the element $p \in R$ is a prime element if and only if R/pR is an integral domain".

- The above proposition gives us a very easy way to construct new fields, which we will explore shortly.

4.2.4 The Chinese Remainder Theorem

- Another foundational result of arithmetic in $\mathbb{Z}$ was the Chinese Remainder Theorem. This result generalizes to arbitrary Euclidean domains, with essentially the same statement.

- We first start with the analogous proposition on solving a single linear congruence.

- Proposition (Linear Congruences): Let R be a Euclidean domain, with $a, b \in R$, and let d any gcd of a and r. Then the equation $ax \equiv b \pmod{r}$ has a solution for $x \in R$ if and only if $d|b$. In this case, if $a = a'd$, $b = b'd$, and $r = r'd$, then $ax \equiv b \pmod{r}$ is equivalent to $a'x \equiv b' \pmod{r'}$ and the solution is $x \equiv (a')^{-1}b' \pmod{r'}$.

  - The proof is the same as over $\mathbb{Z}$.

- Proof: If x is a solution to the congruence $ax \equiv b \pmod{r}$, then there exists an $s \in R$ with $ax - rs = b$. Then since d divides the left-hand side, it must divide b.
- Now if we set $a' = a/d$, $b' = b/d$, and $r' = r/d$, our original equation becomes $a'dx \equiv b'd \pmod{r'd}$.
- Solving this equation is equivalent to solving $a'x \equiv b' \pmod{r'}$, by one of our properties of congruences.
- But since a' and r' are relatively prime, a' is a unit modulo r', so we can simply multiply by its inverse to obtain $x \equiv b' \cdot (a')^{-1} \pmod{r'}$.

- Example: Solve the congruence $(7+i)x \equiv 3 - i$ modulo $8 - 9i$ in $\mathbb{Z}[i]$.
  - Using the Euclidean algorithm we can verify that $7+i$ and $8-9i$ are relatively prime:

$$\begin{aligned} 8 - 9i &= (1-i)(7+i) + (-3i) \\ 7 + i &= (2i)(-3i) + (1+i) \\ -3i &= (-2-2i)(1+i) + i \\ 1 + i &= (1-i)(i) \end{aligned}$$

    and so i, and hence 1, is a gcd.
  - By solving for the remainders we can write 1 as a linear combination explicitly as $1 = (11-i)(7+i) + (-4-5i)(8-9i)$. Hence the inverse of $7+i$ modulo $8-9i$ is $11-i$.
  - Multiplying both sides of the original congruence by $11-i$ yields $x \equiv (11-i)(7+i)x \equiv (11-i)(3-i) \equiv 3+i \pmod{8-9i}$, so the solution is $\boxed{x \equiv 3 + i \pmod{8-9i}}$.

- As over $\mathbb{Z}$, the above proposition converts a problem of solving a general system of congruences in the variable x to a system of the form $x \equiv a_i \pmod{r_i}$.

- Theorem (Chinese Remainder Theorem): Let R be a Euclidean domain and $r_1, r_2, \ldots, r_k$ be pairwise relatively prime elements of R, and $a_1, a_2, \ldots, a_k$ be arbitrary elements of R. Then the system

$$\begin{aligned} x &\equiv a_1 \pmod{r_1} \\ x &\equiv a_2 \pmod{r_2} \\ &\vdots \\ x &\equiv a_k \pmod{r_k} \end{aligned}$$

  has a solution $x_0 \in R$. Furthermore, x is unique modulo $r_1 r_2 \cdots r_k$, and the general solution is precisely the residue class of x_0 modulo $r_1 r_2 \cdots r_k$.
  - The proof is the same as over $\mathbb{Z}$.
  - Proof: Since we may repeatedly convert two congruences into a single one until we are done, by induction it suffices to prove the result for two congruences

$$\begin{aligned} x &\equiv a_1 \pmod{r_1} \\ x &\equiv a_2 \pmod{r_2}. \end{aligned}$$

  - For existence, the first congruence implies $x = a_1 + kr_1$ for some $k \in R$; plugging into the second equation then yields $a_1 + kr_1 \equiv a_2 \pmod{r_2}$. Rearranging yields $kr_1 \equiv (a_2 - a_1) \pmod{r_2}$. Since by hypothesis r_1 and r_2 are relatively prime, by our proposition above we see that this congruence has a unique solution for k modulo r_2, and hence a solution for x.
  - For uniqueness, suppose x and y are both solutions. Then $x - y$ is 0 modulo r_1 and 0 modulo r_2, meaning that $r_1 | (x-y)$ and $r_2 | (x-y)$. But since r_1 and r_2 are relatively prime, their product must therefore divide $x - y$, meaning that x is unique modulo $r_1 r_2$. Finally, it is obvious that any other element of R congruent to x modulo $r_1 r_2$ also satisfies the system.

- Example: In $R = \mathbb{C}[x]$, solve the system $q(x) \equiv 1 \pmod{x-1}$, $q(x) \equiv 3 \pmod{x}$.

- Since $x-1$ and x are relatively prime polynomials, by the Chinese Remainder Theorem all we have to do is find one polynomial satisfying the system.
- If we take the solution $q(x) = 3 + ax$ to the second equation and plug it into the first equation, we must solve $3 + ax \equiv 1 \pmod{x-1}$.
- Since $3 + ax \equiv (3 + a) \bmod (x-1)$, we can take $a = -2$.
- Hence the polynomial $q(x) = 3 - 2x$ is a solution to the system. The general solution is therefore $\boxed{3 - 2x + x(x-1) \cdot s(x)}$ for an arbitrary polynomial $s(x) \in R$. Equivalently, the solution is $\boxed{q(x) \equiv 3 - 2x \pmod{x^2 - x}}$.

4.2.5 Orders, Euler's Theorem, Fermat's Little Theorem

- We can also study powers in R/rR in the same way as in $\mathbb{Z}/m\mathbb{Z}$, with the only caveat being that some elements may not have a finite order:
- Definition: If R is a commutative ring with 1 and u is a unit of R, then the smallest $k > 0$ such that $u^k \equiv 1 \pmod{m}$ is called the order of u. (If there exists no such k, then we say u has infinite order.)
  - Example: The element -1 has order 2 in $\mathbb{Z}$ (and also in $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$), and the element i has order 4 in $\mathbb{Z}[i]$ and in $\mathbb{C}$.
  - Example: The element 2 does not have finite order in $\mathbb{R}$, since no positive power of 2 is equal to 1.
- All of our properties of order hold in general commutative rings with 1:
- Proposition (Properties of Orders): Suppose R is a commutative ring with 1 and u is a unit in R.
  1. If $u^n \equiv 1 \pmod{m}$ for some $n > 0$, then the order of u is finite and 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.
  3. If $u^n \equiv 1 \pmod{m}$ and $u^{n/p} \not\equiv 1 \pmod{m}$ for any prime divisor p of n, then u has order n.
  4. If u has order k and w has order l, where k and l are relatively prime, then uw has order kl.
     - Proof: The proofs are the same as in $\mathbb{Z}/m\mathbb{Z}$.
- One of our foundational results in $\mathbb{Z}/m\mathbb{Z}$ was Euler's theorem. There is a natural generalization of the Euler φ-function and of Euler's theorem that holds in the case where there are finitely many units in R/rR.
- Theorem (Generalization of Euler's Theorem): If R is a commutative ring with 1 and $r \in R$, let $\varphi(r)$ denote the number of units in the ring R/rR, assuming this number is finite. Then if a is any unit in R/rR, we have $a^{\varphi(r)} \equiv 1 \pmod{r}$.
  - The proof is the same as over $\mathbb{Z}/m\mathbb{Z}$: the point is that if a is a unit and $u_1, \cdots, u_k$ are the units in R, then the elements $au_1, \cdots, au_k$ are the same as $u_1, \cdots, u_k$, just in a different order.
  - Proof: Let the set of all units in R/rR be $\overline{u_1}$, $\overline{u_2}$, ... , $\overline{u_{\varphi(r)}}$, and consider the elements $\overline{a \cdot u_1}$, $\overline{a \cdot u_2}$, ..., $\overline{a \cdot u_{\varphi(r)}}$ in R/rR: we claim that they are simply the elements $\overline{u_1}$, $\overline{u_2}$, ... , $\overline{u_{\varphi(r)}}$ again (possibly in a different order).
  - Since there are $\varphi(r)$ elements listed and they are all still units, it is enough to verify that they are all distinct.
  - So suppose $a \cdot u_i \equiv a \cdot u_j \pmod{r}$. Since a is a unit, multiply by a^{-1}: this gives $u_i \equiv u_j \pmod{r}$, but this forces $i = j$.
  - Hence modulo r, the elements $\overline{a \cdot u_1}$, $\overline{a \cdot u_2}$, ..., $\overline{a \cdot u_{\varphi(r)}}$ are simply $\overline{u_1}$, $\overline{u_2}$, ... , $\overline{u_{\varphi(r)}}$ in some order.
  - Therefore we have $(a \cdot u_1)(a \cdot u_2) \cdots (a \cdot u_{\varphi(r)}) \equiv u_1 \cdot u_2 \cdots u_{\varphi(r)} \pmod{r}$, and so cancelling $u_1 \cdot u_2 \cdots u_{\varphi(r)}$ from both sides yields $a^{\varphi(r)} \equiv 1 \pmod{r}$ as desired.
- Although this is the reverse of our approach over $\mathbb{Z}$, we can obtain Fermat's little theorem quite easily using Euler's theorem.

- Corollary (Generalization of Fermat's Little Theorem): If R is a Euclidean domain and $p \in R$ is a prime element, and the number of elements in R/pR is n, then $a^n \equiv a \pmod{p}$ for every $a \in R$.
    - Proof: Since R/pR is a field, the only nonunit is zero, so $\varphi(p) = n - 1$.
    - By the generalization of Euler's theorem, we know that $a^{\varphi(p)} \equiv 1 \pmod{p}$ for every a that is a unit modulo p, so $a^n = a^{\varphi(p)+1} \equiv a \pmod{p}$ for such a.
    - Since $a^n \equiv a \pmod{p}$ is also true when $p|a$, we see that it holds for every $a \in R$.

4.3 Arithmetic in $F[x]$

- In this section, we use all of the ring-theoretic machinery we have developed to study the arithmetic of the polynomial ring $F[x]$.
    - We will first discuss polynomials as functions and use the results to give ways to determine when polynomials of small degree are irreducible.
    - Then we will discuss some of the applications of modular arithmetic in this ring to the construction of finite fields, and (in particular) establish the analogue of the Prime Number Theorem in $\mathbb{F}_p[x]$.
    - We will also use the arithmetic of $F[x]$ to establish the existence of primitive roots in finite fields, and also to characterize the moduli m for which there exists a primitive root.

4.3.1 Polynomial Functions, Roots of Polynomials

- In elementary algebra, polynomials are examples of functions. We would like to extend this idea of "plugging values in" to a general polynomial in $F[x]$, because this allows us to glean some information about potential factorizations.
- Definition: If F is a field and $p = a_0 + a_1 x + \cdots + a_n x^n$ is an element of $F[x]$, for any $r \in F$ we define the value $p(r)$ to be the element $a_0 + a_1 r + \cdots + a_n r^n \in F$.
    - It is straightforward to see from the definition that if p and q are any polynomials in $F[x]$ and r is any element of F, then $(p+q)(r) = p(r) + q(r)$ and $(pq)(r) = p(r)q(r)$. Thus, evaluation at an element of F respects the addition and multiplication structure of the polynomial ring.
    - Example: If $p = 1 + x^2$ in $\mathbb{C}[x]$, then $p(1) = 1 + 1^2 = 2$, and $p(i) = 1 + i^2 = 0$.
    - Example: If $p = 1 + x^2$ in $\mathbb{F}_5[x]$, then $p(0) = 1$, $p(1) = 2$, $p(2) = 0$, $p(3) = 0$, and $p(4) = 2$.
    - In this way, we can view a polynomial $p \in F[x]$ as a function $p : F \to F$, where $p(r) = a_0 + a_1 r + \cdots + a_n r^n$.
    - Warning: The "traditional" polynomial notation $p(x)$ is somewhat ambiguous: we may be considering $p(x)$ as a ring element in $F[x]$ (in which case "x" represents an indeterminate), or we may be viewing it as a function from F to F (in which case "x" represents the variable of the function).
- Example: If $p = x^2 + x$ in $\mathbb{F}_2[x]$, observe that $p(0) = p(1) = 0$.
    - Thus, although p is not the zero polynomial in $\mathbb{F}_2[x]$ (since it has degree 2), as a function from $\mathbb{F}_2$ to $\mathbb{F}_2$ it is the identically zero function!
    - More generally, if F is any finite field with elements $r_1, r_2, \ldots, r_n$, then the polynomial $p(x) = (x - r_1)(x - r_2) \cdots (x - r_n)$ is the identically zero function from F to F.
    - Thus, in general, we cannot always uniquely specify a polynomial $p \in F[x]$ by describing its behavior as a function $p : F \to F$.
- To begin our study of polynomial functions, we start with a pair of observations that are likely familiar from elementary algebra:
- Proposition (Remainder/Factor Theorem): Let F be a field. If $p \in F[x]$ is a polynomial and $r \in F$, then the remainder upon dividing $p(x)$ by $x - r$ is $p(r)$. In particular, $x - r$ divides $p(x)$ if and only if $p(r) = 0$. (In this case we say r is a zero or a root of $p(x)$.)

- Proof: Suppose $p(x) = a_0 + a_1 x + \cdots + a_n x^n$. Observe first that $(x^k - r^k) = (x - r)(x^{k-1} + x^{k-2}r + \cdots + xr^{k-2} + r^{k-1})$, so in particular, $x - r$ divides $x^k - r^k$ for all k.
- Now we simply write $p(x) - p(r) = \sum_{k=0}^{n} a_k(x^k - r^k)$, and since $x - r$ divides each term in the sum, it divides $p(x) - p(r)$.
- Since $p(r)$ is a constant, it is therefore the remainder after dividing $p(x)$ by $x - r$. The other statement is immediate from the uniqueness of the remainder in the division algorithm.

- We can also bound the number of roots that a polynomial can have:
- Proposition (Number of Roots): Let F be a field. If $p \in F[x]$ is a polynomial of degree d, then p has at most d distinct roots in F.
  - Proof: We induct on the degree d. For $d = 1$, the polynomial is of the form $a_0 + a_1 x$ for $a_1 \neq 0$, which has exactly one root, namely $-a_0/a_1$.
  - Now suppose the result holds for all polynomials of degree $\leq d$ and let p be a polynomial of degree $d+1$.
  - If p has no zeroes we are obviously done, so suppose otherwise and let $p(r) = 0$. We can then factor to write $p(x) = (x - r)q(x)$ for some polynomial $q(x)$ of degree d.
  - By the induction hypothesis, $q(x)$ has at most d roots: then $p(x)$ has at most $d + 1$ roots, because $(a - r)q(a) = 0$ only when $a = r$ or $q(a) = 0$ (since F is a field).
- The above results, while seemingly obvious, can fail spectacularly if the coefficient ring is not a field. Here are some especially distressing examples:
  - The quadratic polynomial $q(x) = x^2 - 1$ visibly has four roots modulo 8, namely $x = 1, 3, 5, 7$. Furthermore, $q(x)$ can be factored in two different ways: as $(x - 1)(x - 7)$ and as $(x - 3)(x - 5)$.
  - The linear polynomial $q(x) = x$, despite having degree 1, is not irreducible modulo 6: it can be written as the product $(2x + 3)(3x + 2)$. Furthermore, $q(x) = x$ has one zero (namely $x = 0$), even though its two factors $2x + 3$ and $3x + 2$ each have no zeroes modulo 6.
- In general, it is not easy to determine when an arbitrary polynomial is irreducible. If the degree is small, however, this task can be done by examining all possible factorizations. The following result is frequently useful:
- Proposition (Polynomials of Small Degree): If F is a field and $q(x) \in F[x]$ has degree 2 or 3 and has no zeroes in F, then $q(x)$ is irreducible.
  - Proof: If $q(x) = a(x)b(x)$, taking degrees shows $3 = \deg(q) = \deg(a) + \deg(b)$. Since a and b both have positive degree, one of them must have degree 1. Then its root is also a root of $q(x)$. Taking the contrapositive gives the desired statement.
  - Example: Over $\mathbb{R}$, the polynomial $x^2 + x + 11$ has no roots (since it is always positive), so it is irreducible.
  - Example: Over $\mathbb{F}_5$, the polynomial $q(x) = x^3 + x + 1$ has no roots, since $q(0) = 1$, $q(1) = 3$, $q(2) = 1$, $q(3) = 1$, and $q(4) = 4$.
- For polynomials of larger degree, determining irreducibility can be a much more difficult task. For certain particular fields, we can say more about the structure of the irreducible polynomials.
- Theorem (Fundamental Theorem of Algebra): Every polynomial of positive degree in $\mathbb{C}[x]$ has at least one root. Therefore, the irreducible polynomials in $\mathbb{C}[x]$ are precisely the polynomials of degree 1, and so every polynomial in $\mathbb{C}[x]$ factors into a product of degree-1 polynomials.
  - The first statement of this theorem is a standard result from analysis over the complex numbers, and we take it for granted.
  - To deduce the second statement from the first, observe that if $p(x)$ is any complex polynomial of degree larger than 1, then by assumption it has at least one root r in $\mathbb{C}$, so we can write $p(x) = (x - r)q(x)$ for some other polynomial $q(x)$: then p is reducible.

- Therefore, the irreducible polynomials in $\mathbb{C}[x]$ are precisely the polynomials of degree 1. The final statement follows from the characterization of irreducible polynomials, because every polynomial is a product of irreducibles.

- Another property that we can fruitfully study in a general field is the presence of repeated factors (when we factor a polynomial into irreducibles).
  - Example: Over $\mathbb{C}$, the polynomial $x^3 + x^2 - x - 1$ factors into irreducibles as $(x-1)^2(x+1)$, which has the repeated factor $x - 1$.
  - Example: Over $\mathbb{F}_2$, the polynomial $x^4 + x^2 + 1$ factors into irreducibles as $(x^2 + x + 1)^2$, which has the repeated factor $x^2 + x + 1$.

- As a first goal, we can give a necessary condition for when a polynomial has repeated roots.
  - Recall from calculus that we can test whether a polynomial has a "double root" at r by testing whether $q(r) = q'(r) = 0$. By the factor theorem, this is equivalent to saying that q and q' are both divisible by $x - r$.
  - We can formulate a similar test over any field, since we may give a purely algebraic definition of the derivative:

- Definition: If $q(x) = \sum_{k=0}^{n} a_k x^k$ is a polynomial in $F[x]$, its derivative is the polynomial $q'(x) = \sum_{k=0}^{n} k a_k x^{k-1}$.
  - Example: In $\mathbb{C}[x]$, the derivative of $x^6 - 4x^2 + x$ is $6x^5 - 8x + 1$.
  - Example: In $\mathbb{F}_p[x]$, the derivative of $x^{p^2} - x$ is $p^2 x^{p^2-1} - 1 = -1$. Notice here that although the degree of the original polynomial is p^2, the degree of its derivative is 0.
  - It is a straightforward calculation to verify that the standard differentiation rules apply: $(f+g)'(x) = f'(x) + g'(x)$ and $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$. (For the product rule, the easiest method is to check it for products of monomials and then apply the distributive law, since both sides are additive.)

- Proposition (Repeated Factors): Let F be a field and $q \in F[x]$. Then r is a repeated root of q if and only if $q(r) = q'(r) = 0$. More generally, q has a repeated factor if and only if q and q' are not relatively prime.
  - Proof: First suppose that $q(x)$ has a repeated root r: then $q(x) = (x-r)^2 s(x)$ for some $s(x) \in F[x]$.
  - Taking the derivative yields $q'(x) = 2(x-r)s(x) + (x-r)^2 s'(x) = (x-r) \cdot [2s(x) + (x-r)s'(x)]$. Thus, q' is also divisible by $x - r$ in $F[x]$, so by the factor theorem, we conclude that $q(r) = q'(r) = 0$.
  - Conversely, if $q(r) = q'(r) = 0$, then by the factor theorem $x - r$ divides $q(x)$, so we may write $q(x) = (x-r)a(x)$. Then by the product rule we see that $q'(x) = a(x) + (x-r)a'(x)$, so $q'(r) = a(r)$. Thus $a(r) = 0$ and so $x - r$ divides $a(x)$: then $q(x)$ is divisible by $(x-r)^2$ so r is a repeated root.
  - For the second statement, any root[3] of a common factor of q and q' is a multiple root (by the above) and conversely any repeated root of q will yield a nontrivial common factor of q and q' in $F[x]$.

- Since we can efficiently compute the gcd of $q(x)$ and $q'(x)$ using the Euclidean algorithm in $F[x]$, we can quickly determine if a given polynomial has a repeated factor.

- Example: Determine whether $q(x) = x^4 + 3x^3 + 3x^2 + 3x + 1$ has a repeated factor in $\mathbb{F}_5[x]$.
  - We have $q'(x) = 4x^3 + 4x^2 + x + 3$.
  - Now we perform the Euclidean algorithm: this yields

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

    and so since $2x^2 + 3x + 2$ is a greatest common divisor (it is associate to monic polynomial $x^2 + 4x + 1$) we see that $q(x)$ $\boxed{\text{has a repeated factor}}$.
  - Indeed, if we divide $q(x)$ by $x^2 + 4x + 1$, we will see that $q(x) = (x^2 + 4x + 1)^2$.

[3] We note here that the common factor may not have any roots in F, in which case one must (in general) instead work in a field extension K/F in which this polynomial does have a root. Such an extension can always be proven to exist, as we will see later.

4.3.2 Finite Fields

- We can fruitfully apply our results to the case where $F = \mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is a finite field with p elements:
- Theorem (Finite Fields): If $q(x) \in \mathbb{F}_p[x]$ is an irreducible polynomial of degree d, then the ring R/qR is a finite field with p^d elements.
  - Proof: We simply invoke our previous results: the residue classes in the ring R/qR are given by the polynomials in $\mathbb{F}_p[x]$ of degree $\leq d-1$.
  - Such a polynomial has the form $a_0 + a_1 x + \cdots + a_{d-1}x^{d-1}$, where the coefficients a_i are arbitrary elements of $\mathbb{F}_p$. There are clearly p^d such polynomials.
  - Now by our earlier results, we know that if $q(x)$ is irreducible, R/qR is an integral domain. Since it is also finite, it is a field.
- Example: Show that the ring R/qR, where $R = \mathbb{F}_2[x]$ and $q(x) = x^2 + x + 1$, is a field with 4 elements.
  - This follows because $x^2 + x + 1$ is irreducible modulo 2: if it had a nontrivial factorization, then since it is a polynomial of degree 2, it would necessarily have a root (which it does not).
  - We showed this fact explicitly earlier when we wrote out the addition and multiplication tables for this field.
- Example: Show that the ring R/qR, where $R = \mathbb{F}_3[x]$ and $q(x) = x^2 + 1$, is a field with 9 elements.
  - This follows because $x^2 + x + 1$ is irreducible modulo 3, since, if it had a nontrivial factorization, then since it is a polynomial of degree 3, it would necessarily have a root (which it does not).
  - Explicitly, the nine elements of this field are 0, 1, 2, x, $x+1$, $x+2$, $2x$, $2x+1$, and $2x+2$. Addition is taken with coefficients modulo 3, and multiplication is performed under the convention that $x^2 + 1 = 0$ (i.e., $x^2 = 2$, since coefficients are taken modulo 3).
  - Here is the multiplication table for this field:

$\cdot$	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
2	0	2	1	$2x$	$2x+2$	$2x+1$	x	$x+2$	$x+1$
x	0	x	$2x$	2	$x+2$	$2x+2$	1	$x+1$	$2x+1$
$x+1$	0	$x+1$	$2x+2$	$x+2$	$2x$	1	$2x+1$	2	x
$x+2$	0	$x+2$	$2x+1$	$2x+2$	1	x	$x+1$	$2x$	2
$2x$	0	$2x$	x	1	$2x+1$	$x+1$	$2x+2$	$2x+1$	$d+2$
$2x+1$	0	$2x+1$	$x+2$	$x+1$	2	$2x$	$2x+2$	x	1
$2x+2$	0	$2x+2$	$x+1$	$2x+1$	x	2	$x+2$	1	$2x$

  - One can check, for example, that $x+1$, $x+2$, $2x+1$, and $2x+2$ are all primitive roots in this field.
- Example: Construct a finite field with 8 elements.
  - From our discussion, since $8 = 2^3$, such a field can be obtained as R/qR where $R = \mathbb{F}_2[x]$ and q is an irreducible polynomial in R of degree 3.
  - It is easy to see that $q(x) = x^3 + x + 1$ is irreducible in $\mathbb{F}_2[x]$ since it has no roots, so R/qR is a finite field with 8 elements.
- We can now use Fermat's little theorem in these finite fields to extract interesting and useful information.
  - To start, observe that by (the original) Fermat's little theorem, $a^p \equiv a \pmod{p}$. Thus, if $q(x) = x^p - x$, then $q(a) = 0$ for every $a \in \mathbb{F}_p$.
  - In other words, this polynomial $x^p - x$ has the rather strange property that its value is always zero, yet it is not the zero polynomial.
- Proposition (Factorization of $x^p - x$): The factorization of $x^p - x$ in $\mathbb{F}_p[x]$ is $x^p - x = \prod_{a \in \mathbb{F}_p}(x - a)$.

- Proof: As noted above, $q(x) = x^p - x$ is such that $q(a) = 0$ for every $a \in \mathbb{F}_p$.
- Hence, $x - a$ is a divisor of $q(x)$ for every $a \in \mathbb{F}_p$.
- However, because this polynomial has at most p roots, and we have exhibited p roots, the factorization of $q(x)$ must be $q(x) = \prod_{a \in \mathbb{F}_p}(x - a)$, since the leading terms agree.

- Another immediate application of this factorization is an easy proof of Wilson's Theorem.
  - By dividing through by x, we see that $x^{p-1} - 1 = \prod_{a \in \mathbb{F}_p, a \neq 0}(x - a)$.
  - Now examine the constant term of the product: it is $(-1)^{p-1} \prod_{a \in \mathbb{F}_p, a \neq 0}(a) = (-1)^{p-1} \cdot (p-1)!$.
  - But (modulo p) the constant term is also equal to -1, so we deduce $(p-1)! \equiv (-1)^{p-2} \equiv -1 \pmod{p}$.

- As we observed above, the polynomial $x^p - x$ has a nice factorization in $\mathbb{F}_p[x]$. Let us now consider the factorization of the polynomial $x^{p^n} - x$ in $\mathbb{F}_p[x]$.
  - 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 list of irreducible factors of $x^{25} - x$ is 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.
  - In fact, it seems that the factorization of $x^{p^n} - x$ over $\mathbb{F}_p$ contains all of the irreducible polynomials of degree n, or of degree dividing n.

- Theorem (Factorization of $x^{p^n} - x$): 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 there are 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, we have $q'(x) = p^n x^{p^n - 1} - 1 = -1$, so $q(x)$ and $q'(x)$ are relatively prime. Thus, by our earlier results, we know that $q(x)$ has no repeated irreducible factors.
  - Before starting the rest of the proof, we first show a simple lemma:
  - Lemma: If p is any prime, the gcd of $p^n - 1$ and $p^d - 1$ is $p^{\gcd(n,d)} - 1$.
    - 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, now suppose that $s(x) \in \mathbb{F}_p[x]$ is an irreducible polynomial of degree d, where $n = ad$.
  - We know that R/sR is a finite field F having p^d elements, so by Euler's theorem in F, 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, R/sR is a finite field F 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 $a \in F$ 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 $\mathbb{F}_p$.

- As a corollary, the above theorem allows us to count the number of monic irreducible polynomials in $\mathbb{F}_p[x]$ of any particular degree n.
  - Let $f_p(n)$ be the number of monic irreducible polynomials of exact degree n in $\mathbb{F}_p[x]$.
  - The theorem above 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, the formula says that there are 2 irreducible polynomials of degree 3 over $\mathbb{F}_2$, which there are: $x^3 + x^2 + 1$ and $x^3 + x + 1$.
  - In fact, we can essentially write down a general formula.

- 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, we have

$$\begin{aligned} \sum_{d|n} \mu(d) g(n/d) &= \sum_{d|n} \mu(d) \sum_{d'|(n/d)} f(d') \\ &= \sum_{dd'|n} \mu(d) f(d') \\ &= \sum_{d'|n} f(d') \sum_{d|(n/d')} \mu(d) \end{aligned}$$

  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: If $f_p(n)$ is the number of monic irreducible polynomials of degree n in $\mathbb{F}_p[x]$, then $f_p(n) = \frac{1}{n}\sum_{d|n} p^{n/d}\mu(d)$.
    - 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 $p^{n/2}$.
- This has the following interesting reinterpretation: let X be the number of polynomials in $\mathbb{F}_p[x]$ of degree less than n.
    - Clearly, $X = p^n$.
    - Now we ask: of all 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: in fact, it is the analogue of the Prime Number Theorem for the ring $\mathbb{F}_p[x]$.
- It is also fairly easy to show using the formula that $f_p(n) > 0$ for every prime p and every integer $n \geq 1$. As we showed earlier, if $q(x)$ is an irreducible polynomial of degree n in $R = \mathbb{F}_p[x]$, then R/qR is a finite field of size p^n. Thus, we also obtain the following:
- Corollary: For any prime p and any n, there exists a finite field having p^n elements.
    - Remark: It can be shown that a finite field must have prime-power order[4], so this result completely characterizes the number of elements that a finite field can have.

4.3.3 Primitive Roots

- We discussed primitive roots previously, but did not categorize when they do or do not exist modulo m. We will extend our viewpoint slightly and treat primitive roots in arbitrary rings:
- Definition: If R is a commutative ring with 1 having finitely many units, an element $u \in R$ is a primitive root if every unit of R is some power of u.
    - More explicitly, if there are n units in R, then an element is a primitive root precisely when its order is n.
- Our first goal is to prove that every finite field has a primitive root. To do so we require the following preliminary fact:
- Proposition: Let R be a commutative ring with 1 having finitely many units. If M is the maximal order among all units in R, then the order of every unit divides M.
    - Proof: Suppose u has order M, and let w be any other unit, of order k.
    - Suppose 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.

[4] To summarize: if K is a finite field, if we let K' be the subfield of K generated by the element 1 (in other words, the subfield whose elements are 0, 1, $1+1$, $1+1+1$, ...), it can be shown that K' has a prime number of elements p, and that K is a vector space over K'. Then because every vector space has a basis, if we select a basis with d elements for K as a vector space over K', then by counting the possible linear combinations of the basis elements we see that the number of elements in K is p^d, which is a prime power.

- Since these two orders are relatively prime, the element $u^{q^f} \cdot w^{k/q^e}$ has order $M \cdot q^{f-e}$, which is a contradiction because this is larger than M.
- Remark (for those who like group theory): This result actually holds in any abelian group, with the same proof: if M is the maximal order among all elements of a finite abelian group, then the order of every element divides M.

- Theorem (Primitive Roots in Finite Fields): If F is a finite field, then F has a primitive root.
  - Our proof is nonconstructive: we will show the existence of a primitive root without explicitly finding one.
  - Proof: Suppose M is the maximal order among all units in F. Then by the finite-field version of Euler's theorem, we know that $M \leq |F| - 1$, since $a^{|F|-1} = 1$ in F for every unit $a \in F$.
  - By the above proposition, all units in F then have order dividing M, so the polynomial $x^M - 1$ has $|F| - 1$ roots in F.
  - But this is impossible unless $M \geq |F| - 1$, since a polynomial of degree M can only have at most M roots in F.
  - Hence we conclude $M = |F| - 1$, meaning that some element has order $|F| - 1$: this element is a primitive root.

- By setting $F = \mathbb{Z}/p\mathbb{Z}$, we obtain the existence of a primitive root modulo p. Using this, it turns out that we can easily construct a primitive root modulo p^2:

- Proposition (Primitive Roots Modulo p^2): If a is a primitive root modulo p for p an odd prime, then a is a primitive root modulo p^2 if $a^{p-1} \not\equiv 1 \pmod{p^2}$. In the event that $a^{p-1} \equiv 1 \pmod{p^2}$, then $a + p$ is a primitive root modulo p^2.
  - Proof: Since a is a primitive root modulo p, if the order of a mod p^2 is r, then since $a^r \equiv 1 \pmod{p^2}$ certainly implies $a^r \equiv 1 \pmod{p}$, we see that $p - 1$ divides r.
  - Since $\varphi(p^2) = p(p-1)$, there are two possibilities: the order of a modulo p^2 is $p - 1$ or it is $p(p-1)$.
  - The order of a modulo p^2 will be $p - 1$ if and only if $a^{p-1} \equiv 1 \pmod{p^2}$. This gives the first statement.
  - For the second statement, suppose that $a^{p-1} \equiv 1 \pmod{p^2}$.
  - The binomial theorem implies $(a+p)^{p-1} = a^{p-1} + (p-1)p \cdot a^{p-2} + p^2 \cdot [\text{other terms}]$, which is simply $a^{p-1} - p\,a^{p-2} \pmod{p^2}$.
  - Since $a^{p-1} \equiv 1 \pmod{p^2}$, we see that $a^{p-2} - p\,a^{p-2}$ cannot be equivalent to 1 mod p^2, because $p\,a^{p-2}$ is not divisible by p^2. So $(a+p)^{p-1} \not\equiv 1 \pmod{p^2}$, so by the earlier argument $a + p$ is a primitive root modulo p^2.

- Example: Find a primitive root modulo 11^2.
  - First we show that that 2 is a primitive root modulo 11: since the order of 2 must divide $\varphi(11) = 10$, and we see that $2^2 \not\equiv 1 \pmod{11}$ and $2^5 \not\equiv 1 \pmod{11}$, the order divides neither 2 nor 5, hence must be 10.
  - Now we can easily compute $2^{10} = 1024 \equiv 56 \pmod{11^2}$, so the proposition above assures us that 2 is also a primitive root modulo 11^2.

- Now we examine primitive roots modulo p^d for $d > 2$; it turns out that these are essentially the same as primitive roots modulo p^2:

- Proposition (Primitive Roots Modulo p^d): If a is a primitive root modulo p^2 for p an odd prime, then a is a primitive root modulo p^d for all $d \geq 2$.
  - Proof: We show this by induction on d: the base case $d = 2$ is vacuous.
  - Now suppose that a is a primitive root modulo p^d and that it has order r modulo p^{d+1}: thus, $a^r \equiv 1 \pmod{p^{d+1}}$. Note that Euler's theorem implies that r divides $\varphi(p^{d+1}) = p^d(p-1)$.

- Reducing modulo p^d shows $a^r \equiv 1 \pmod{p^d}$, so since a is a primitive root modulo p^d we see that r is divisible by $\varphi(p^d) = p^{d-1}(p-1)$.
- Thus, the only possibilities are $r = p^{d-1}(p-1)$ and $r = p^d(p-1)$: we just need to eliminate the first possibility.
- By Euler's theorem, $a^{p-1} \equiv 1 \pmod{p}$ so we can write $a^{p-1} = 1 + kp$ for some integer k.
- Then, since a is a primitive root modulo p^2, we also know that k is not divisible by p (as otherwise a would have order $p-1$ modulo p^2).
- Expanding with the binomial theorem yields $(a^{p-1})^{p^{d-1}} = (1+kp)^{p^{d-1}} = 1 + p^{d-1} \cdot kp + p^{d+1} \cdot [\text{other terms}]$. But this is $\not\equiv 1$ modulo p^{d+1}, since k is not divisible by p.
- Hence $a^{p^{d-1}(p-1)} \not\equiv 1 \pmod{p^{d+1}}$, so a must have order $p^d(p-1) = \varphi(p^{d+1})$, meaning a is in fact a primitive root.

- Example: Find a primitive root modulo 11^{100}.
  - We saw in the previous example that 2 was a primitive root modulo 11^2. Hence by the proposition above, 2 is a primitive root modulo 11^d for any $d \geq 2$ hence (in particular) for $d = 100$.

- Given a primitive root modulo p^d, it is easy to construct a primitive root modulo $2p^d$:

- Proposition: If a is a primitive root modulo p^d for p an odd prime, then a is a primitive root modulo $2p^d$ if a is odd, and $a + p^d$ is a primitive root modulo $2p^d$ if a is even.
  - Proof: If a is odd, then a, a^2, ... , $a^{\varphi(p^d)}$ are all odd and distinct modulo p^d. Hence they all remain invertible modulo $2p^d$, and are clearly still distinct.
  - But since $\varphi(2p^d) = \varphi(p^d)$, the elements a, a^2, ... , $a^{\varphi(p^d)}$ exhaust all of the distinct residue classes modulo $2p^d$, meaning that a is a primitive root.
  - If a is even, then $a + p^d$ is odd, and we can apply the argument above to see $a + p^d$ is a primitive root modulo $2p^d$.

- Example: Find a primitive root modulo $2 \cdot 11^{100}$.
  - From before, we know that 2 is a primitive root modulo 11^{100}. Since 2 is odd, the above corollary implies that $2 + 11^{100}$ is a primitive root modulo $2 \cdot 11^{100}$.

- With all of the above results, we can now finish the characterization of the moduli that have primitive roots:

- Theorem (Primitive Roots Modulo m): There exists a primitive root modulo m if and only if $m = 1, 2, 4$, or $m = p^k$ or $2p^k$ for an odd prime p and some $k \geq 1$.
  - Proof: We have already shown the existence of primitive roots in all of these cases except $m = 1, 2, 4$, but the existence of a primitive root for those moduli is trivial. All we have left to do is show that a primitive root cannot exist for other m.
  - We begin with the observation that if there exists a primitive root r modulo m, then necessarily the congruence $x^2 \equiv 1 \pmod{m}$ has only two solutions modulo m.
    - Suppose $u = r^d$ for some $0 \leq d < \varphi(m)$ is a solution to $u^2 \equiv 1 \pmod{m}$.
    - Then $r^{2d} \equiv 1 \bmod m$, so since r has order $\varphi(m)$ there are only two possibilities for d, namely $d = 0$ and $d = \varphi(m)/2$.
    - Hence, there are only two possible u (which are, indeed, $u = 1$ and $u = -1$).
  - We then see that there cannot exist a primitive root modulo $4p$ for any prime p (including $p = 2$).
    - The congruence $x^2 \equiv 1 \pmod{4p}$ has the four distinct solutions $x \equiv \pm 1$ and $x \equiv \pm(2p-1)$, so by the above there cannot be a primitive root.
  - Similarly, there cannot exist a primitive root modulo pq for any distinct odd primes p and q.
    - By the Chinese Remainder Theorem, there are four solutions to $x^2 \equiv 1 \pmod{pq}$, obtained by solving the congruences $x \equiv \pm 1 \pmod{p}$ and $x \equiv \pm 1 \pmod{q}$ simultaneously.

- We also note that if r is a primitive root modulo m and d divides m, then r is a primitive root modulo d.
  - If the powers of r yield all the invertible residue classes modulo m, then they certainly yield all the invertible residue classes modulo d.
- Therefore: if m is divisible by $4p$ for any prime p, or is divisible by two distinct odd primes, there is no primitive root modulo m. These two cases together encompass everything we needed to show, so we are done.

- For completeness, we restate a result we showed in a previous chapter about the number of primitive roots modulo m:
- Proposition (Number of Primitive Roots): If there exists a primitive root modulo m, then there are precisely $\varphi(\varphi(m))$ primitive roots modulo m.
  - Proof: Suppose that there exists a primitive root u modulo m, whose order is therefore $\varphi(m)$.
  - We know that the invertible residue classes modulo p are represented by $u^1, \ldots, u^{\varphi(m)}$, so it suffices to determine how many of these have order $\varphi(m)$.
  - Since the order of u^k is $\varphi(m)/\gcd(k, \varphi(m))$, we see that u^k is a primitive root if and only if k is relatively prime to $\varphi(m)$.
  - There are $\varphi(\varphi(m))$ such k, so there are $\varphi(\varphi(m))$ primitive roots modulo m.

4.4 Arithmetic in $\mathbb{Z}[i]$

- In this section, we use all of the ring-theoretic machinery we have developed to study the arithmetic of the Gaussian integer ring $\mathbb{Z}[i]$.
  - Our first goal is to study modular arithmetic in this ring.
  - Then we turn our attention to characterizing the irreducible elements in this ring. 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 with the term "prime number" when we refer to rational integers 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, and that this map is multiplicative: $N(zw) = N(z)N(w)$.
- We collect a few basic facts about norms that hold in $\mathbb{Z}[\sqrt{D}]$ for a general D:
- Proposition: If $\alpha \in \mathbb{Z}[\sqrt{D}]$, then α is a unit if and only if $N(\alpha) = \pm 1$. Also, if $N(\alpha) = \pm p$ for a prime p, then α is irreducible.
  - Proof: If α is a unit, say with $\alpha\beta = 1$, then $1 = N(1) = N(\alpha\beta) = N(\alpha)N(\beta)$, so $N(\alpha) = \pm 1$.
  - If $N(\alpha) = \pm 1$, then if $\alpha = a + b\sqrt{D}$, we have $(a - b\sqrt{D})(a + b\sqrt{D}) = N(\alpha) = \pm 1$, so α times $\pm(a - b\sqrt{D})$ is equal to 1, meaning α is a unit.
  - Finally, if $\alpha = \beta\gamma$ and $N(\alpha) = \pm p$, then $N(\beta)N(\gamma) = \pm p$. If p is prime, then one of $N(\beta)$, $N(\gamma)$ must be ± 1, so by the above one of β, γ is a unit.

4.4.1 Residue Classes in $\mathbb{Z}[i]$

- A natural question is: if $\beta \in \mathbb{Z}[i]$ is some arbitrary element, how many residue classes modulo β are there, and is there an easy way to write them down?
  - It might seem as though the division algorithm would give them to us: we proved that for any $\alpha \in \mathbb{Z}[i]$, there exist a $q, r \in \mathbb{Z}[i]$ such that $\alpha = q\beta + r$, and where $N(r) \leq \frac{1}{2}N(\beta)$.

- Thus, the collection of possible remainders r with $N(r) \leq \frac{1}{2}N(\beta)$ certainly give all the residue classes.
- However, the quotient and remainder arising in the division algorithm are not guaranteed to be unique: there can be more than one possible r such that $\alpha \equiv r \pmod{\beta}$ and $N(r) < \frac{1}{2}N(\beta)$.

- It turns out that it is much easier to understand the modular arithmetic in $\mathbb{Z}[i]$ from a geometric point of view.
  - In the complex plane, the Gaussian integers form the set of lattice points, the points whose coordinates are both integers. We can also view Gaussian integers as vectors in this lattice, since the additive structure of $\mathbb{Z}[i]$ agrees with the additive structure of vectors in the plane.

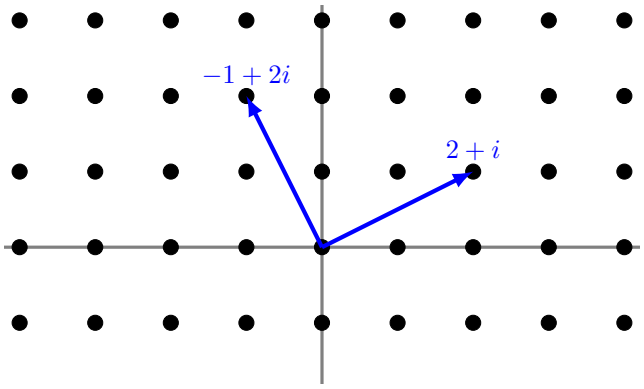

Figure 1: The Gaussian integers as a lattice, and the two vectors $\beta = 2+i$ and $i\beta = -1+2i$.

  - Now consider the multiples of a given Gaussian integer β: every multiple is of the form $(x+iy)\beta = x\beta + y(i\beta)$, so it is an integer linear combination of β and $i\beta$.
  - Thus, drawing all of the $\mathbb{Z}[i]$-multiples of β is the same as drawing all of the vectors that can be obtained by an integer number of "steps" each in the direction of β or $i\beta$, which produces a square tiling of the plane.

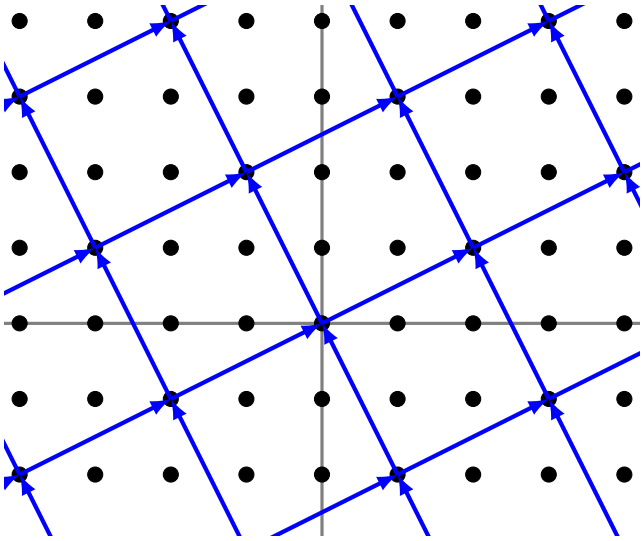

Figure 2: The $\mathbb{Z}[i]$-multiples of $\beta = 2+i$ with marked vectors $\beta = 2+i$ and $i\beta = -1+2i$.

  - Geometrically, two Gaussian integers will be congruent modulo β if and only if they are located in the same position within two different squares.

- Thus, if we take the collection of lattice points inside any one of these squares, it will yield a "fundamental region" for the Gaussian integers modulo β: the elements in the fundamental region will be unique representatives for the residue classes modulo β.

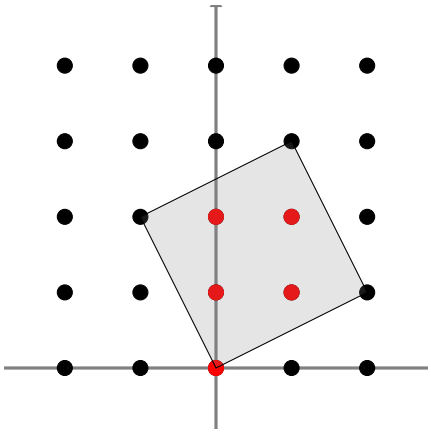

Figure 3: A fundamental region for $\mathbb{Z}[i]$ modulo $\beta = 2 + i$ and a marked set of representatives.

- As shown in the figures, there is a fundamental region for $\mathbb{Z}[i]$ modulo $2 + i$ containing the 5 points 0, i, $2i$, $1 + i$, and $2 + i$.
  - Hence, every element of $\mathbb{Z}[i]$ is congruent modulo $2 + i$ to 0, i, $2i$, $1 + i$, or $2 + i$.
  - We conclude that there are 5 residue classes modulo $2 + i$. (Recall that we showed this earlier using a different approach.)
- Notice that $N(2 + i) = 5$, and there are 5 residue classes modulo $2 + i$. In general, it turns out that there are exactly $N(\beta)$ residue classes modulo β for any nonzero β. We can prove this using (of all things) a theorem from elementary geometry!
- Theorem (Pick's Theorem): If R is a polygon in the plane whose vertices are all lattice points, then the area of R is given by the formula $A = I + \frac{1}{2}B - 1$, where I is the number of lattice points in the interior of R and B is the number of lattice points on the boundary of R.
  - Remark: We say a point of R is a boundary point if it lies on one of the sides of R. We say a point of R is an interior point if it does not lie on one of the sides of R.
  - The result is easiest to see with an example: by drawing a rectangle around the given polygon and subtracting small triangles, one can see that this polygon has area $\frac{17}{2} = 5 + \frac{9}{2} - 1$.

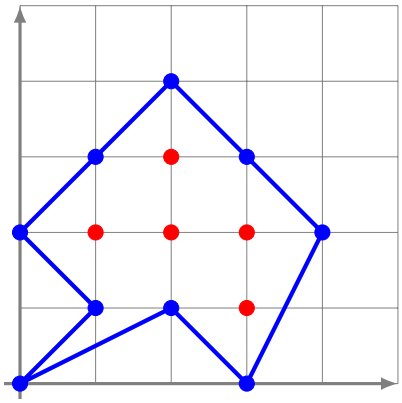

Figure 4: A lattice polygon with 9 boundary points (in blue) and 5 interior points (in red).

- We will omit the full proof[5], since it is not really relevant to our goals.

- We can use Pick's theorem to give an easy computation of the number of residue classes in $\mathbb{Z}[i]$ modulo β:
- Theorem (Number of Residue Classes in $\mathbb{Z}[i]/\beta$): If β is a nonzero Gaussian integer, the number of distinct residue classes in $\mathbb{Z}[i]$ modulo β is equal to $N(\beta)$.
  - Proof: Consider a fundamental region for $\mathbb{Z}[i]$ modulo β.
  - By our geometric arguments above, every Gaussian integer has a unique representative modulo β that lies in the fundamental region, which we can take to be the square whose vertices are 0, β, $i\beta$, and $\beta + i\beta$ in the complex plane.
  - Each interior point of this square yields one residue class.
  - The boundary points of the square come in pairs (on opposite edges of the square) each yielding one residue class, except for the four vertices (0, β, $i\beta$, $\beta + i\beta$) which all lie in the same residue class.
  - Thus, the total number of residue classes is $I + \dfrac{B-4}{2} + 1 = I + \dfrac{1}{2}B - 1$.
  - But by Pick's Theorem, this is precisely the area of the fundamental region. Since this region is a square with side length $|\beta|$, the area is simply $|\beta|^2 = N(\beta)$.

- Thus, to list all of the residue classes modulo $\beta \in \mathbb{Z}[i]$, we need only give a list of $N(\beta)$ inequivalent residue classes, which must therefore be exhaustive.
  - To generate this list, we can draw a fundamental region for $\mathbb{Z}[i]$ modulo β.

- Example: Find representatives for the residue classes modulo $2 + 2i$ in $\mathbb{Z}[i]$.
  - We have $N(2 + 2i) = 8$ so there are 8 residue classes. It is then not hard to verify that

4.4.2 Prime Factorization in $\mathbb{Z}[i]$

- We now turn our attention to factorization in $\mathbb{Z}[i]$.
  - If $\pi \in \mathbb{Z}[i]$, π certainly divides $N(\pi)$. So if π is irreducible in $\mathbb{Z}[i]$, then since irreducibles are prime elements in a Euclidean domain, we conclude that π must divide one of the (integer) prime factors of the integer $N(\pi)$.
  - Therefore, to characterize the irreducible elements of $\mathbb{Z}[i]$, we need to study how primes $p \in \mathbb{Z}$ factor in $\mathbb{Z}[i]$.

- 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).

[5] To summarize: first establish Pick's theorem for rectangles (an easy counting argument), and that it is consistent with taking unions of regions along an edge and also with removing a portion of a region along an edge. Then deduce that it holds for right triangles, then for all triangles, and finally that any polygonal region can be constructed by adding or subtracting triangular regions from rectangles.

- 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 $\equiv 1 \pmod{p}$. We will show later that half of the units modulo p will have $u^{(p-1)/2} \equiv -1 \pmod{p}$, in which 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 n. (Note of course that we can compute $u^{(p-1)/4}$ very efficiently using successive squaring.)

- 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 $4(n_1+1)(n_2+1)\cdots(n_k+1)$.
  - 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 $1 \le i \le k$.
  - 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 16 ways can be found from the different ways of choosing one of $5 \pm 6i$ and multiplying it with $10 \pm 3i$.
  - Explicitly: $(5+6i)(10+3i) = 32+75i$, and $(5+6i)(10-3i) = 68+45i$, so we obtain the sixteen ways of writing 6649 as the sum of two squares as $(\pm 32)^2 + (\pm 75)^2$, $(\pm 68)^2 + (\pm 45)^2$, and the eight other decompositions with the terms interchanged.

- 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.

- As a third corollary of our classification, we obtain another way to construct finite fields: if $p \in \mathbb{Z}$ is a prime congruent to 3 modulo 4, then, for $R=\mathbb{Z}[i]$, we know that R/pR is a field of size $N(p)=p^2$.
  - Using our description of the fundamental region for R/pR, we can see that a set of residue class representatives is given by the elements of the form $a+bi$ for $0 \le a, b \le p-1$, where coefficients are taken modulo p.

- Example: For $p=3$, we obtain the field of order 9 whose elements are 0, 1, 2, i, $1+i$, $2+i$, $2i$, $1+2i$, and $2+2i$, where for example, we have $(1+i)\cdot(2+i)=1+3i \equiv 1 \pmod 3$.
  - Note: Technically, we should put lines over all of the elements to emphasize that they are residue classes, but this would be confusing since the complex conjugate is also denoted the same way.
  - Notice that we constructed another field of order 9 earlier, namely the quotient of $\mathbb{F}_3[x]$ by the irreducible polynomial x^2+1 where, for example, we have $(1+x)\cdot(2+x)=2+3x+x^2 \equiv 1$.
  - As can be verified by trying out a few more examples, the arithmetic in these two fields is exactly the same (simply replace the number i with the variable x).
  - This should not be too surprising: $\mathbb{F}_3[x]$ modulo x^2+1 is obtained from $\mathbb{Z}$ by first declaring that 3 is equal to 0 (to create $\mathbb{F}_3=\mathbb{Z}/3\mathbb{Z}$), and then including a new element $\overline{x}$ whose square is -1.
  - On the other hand, $\mathbb{Z}[i]$ modulo 3 is obtained from $\mathbb{Z}$ by first introducing a new element i whose square is -1, and then declaring that 3 is equal to 0.

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