Contents

1 Integers, Polynomials, and Rings

Our goal in this chapter is to study the structure of polynomials, generalizing the idea of a "polynomial with real coefficients" familiar from elementary algebra. As motivation, we will begin by briefly reviewing the arithmetic of the integers ($\mathbb{Z}$) and the integers modulo m ($\mathbb{Z}/m\mathbb{Z}$), and then describe many of the analogous properties of polynomials, including the division algorithm, the Euclidean algorithm, and factorization into irreducibles. Then we will discuss some additional features unique to polynomials, such as various criteria for irreducibility along with polynomial modular arithmetic. Finally, we give a brief overview of rings and some of their basic properties, and describe how rings yield a generalization of much of our discussion of integers and polynomials.

1.1 The Integers and Modular Arithmetic

- We are all quite familiar with the integers $\mathbb{Z}$, consisting of the natural numbers $\mathbb{N}$ (1, 2, 3, 4, ...), along with their negatives (-1, -2, -3, -4, ...) and zero (0). There are two natural binary arithmetic operations defined on the integers, namely addition ($+$) and multiplication ($\cdot$), along with the unary operation of negation ($-$).
- It is not quite so simple to prove things about the integers without a solid set of properties to work from. For concreteness, one may give a completely explicit axiomatic description of the integers as follows:

- "Definition": The integers are a set $\mathbb{Z}$ along with two (closed) binary[1] operations $+$ and $\cdot$, obeying the following properties[2]:

  [I1] The operation $+$ is associative: $a + (b + c) = (a + b) + c$ for any integers a, b, c.

  [I2] The operation $+$ is commutative: $a + b = b + a$ for any integers a, b.

  [I3] There is an additive identity 0 satisfying $a + 0 = a$ for all integers a.

  [I4] Every integer a has an additive inverse $-a$ satisfying $a + (-a) = 0$.

  [I5] The operation $\cdot$ is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for any integers a, b, c.

  [I6] The operation $\cdot$ is commutative: $a \cdot b = b \cdot a$ for any integers a, b.

  [I7] There is a multiplicative identity $1 \neq 0$ satisfying $1 \cdot a = a$ for all integers a.

  [I8] The operation $\cdot$ distributes over $+$: $a \cdot (b + c) = a \cdot b + a \cdot c$ for any integers a, b, c.

  Furthermore, there is a subset of $\mathbb{Z}$, called $\mathbb{N}$, such that

  [N1] For every $a \in \mathbb{Z}$, precisely one of the following holds: $a \in \mathbb{N}$, $a = 0$, or $(-a) \in \mathbb{N}$.

  [N2] $\mathbb{N}$ is closed under $+$ and $\cdot$: for any $a, b \in \mathbb{N}$, both $a + b$ and $a \cdot b$ are in $\mathbb{N}$.

  [N3] Every nonempty subset S of $\mathbb{N}$ contains a smallest element: that is, an element $x \in S$ such that if $y \in S$, then either $y = x$ or $y - x \in \mathbb{N}$.

- Remark: The axiom (N3) is called the well-ordering principle. It is the axiom that differentiates the integers from other number systems such as the rational numbers or the real numbers (both of which obey all of the other axioms).

  - One may use the axiomatic description of the integers to establish all of the standard properties of arithmetic: for example, we can define the binary operation of subtraction by setting $a - b = a + (-b)$, as well as the order relation "$<$" by saying $a < b$ if and only if $b - a \in \mathbb{N}$. (We define $b > a$ to be the same thing.)
  - Using the axioms, one can then establish results like the following: the elements 0 and 1 are unique, additive inverses are unique, 1 is a positive integer, $a + b = a + c$ implies $b = c$, $0 \cdot a = 0$ for any a, $a \cdot b = a \cdot c$ and $a \neq 0$ implies $b = c$, there are no integers between 0 and 1, and so forth.

- It is incredibly tedious to write proofs relying solely on axiomatic calculations, so from this point we will simply work in standard notation.

1.1.1 Divisibility, GCDs, the Euclidean Algorithm, and Prime Factorization

- So far we have discussed addition, subtraction, and multiplication. Division is trickier, because it is not always possible to divide one integer by another and obtain an integer quotient, so first we analyze the situation when "division" is possible:

- Definition: If $a \neq 0$, we say that a divides b (equivalently, b is divisible by a), written $a|b$, if there is an integer k with $b = ka$.

  - Examples: $2|4$, $(-7)|7$, and $6|0$.

- There are a number of basic properties of divisibility that follow immediately from the definition and properties of arithmetic:

  - If $a|b$, then $a|bc$ for any c.
  - If $a|b$ and $b|c$, then $a|c$.
  - If $a|b$ and $a|c$, then $a|(xb + yc)$ for any x and y.

[1] The definition of a binary operation means that for any two integers a and b, the symbols $a + b$ and $a \cdot b$ are always defined and are integers. Some authors list these properties explicitly as part of their list of axioms.

[2] To be a proper definition, we would also need to establish that there actually is a set with operations obeying these properties, which turns out to be rather difficult. But there are various constructions for $\mathbb{Z}$ using set theory, which we will not detail here.

- If $a|b$ and $b|a$, then $a = b$ or $a = -b$.
- If $a|b$, and $a, b > 0$, then $a \leq b$.
- For any $m \neq 0$, $a|b$ is equivalent to $(ma)|(mb)$.

- If $0 < b < a$ and b does not divide a, we can still attempt to divide a by b to obtain a quotient and remainder: this is a less-explicit version of the long-division algorithm familiar from elementary school. Formally:

- Theorem (Division Algorithm): If a and b are positive integers, then there exist unique integers q and r such that $a = qb + r$ with $0 \leq r < b$. Furthermore, $r = 0$ if and only if $b|a$.

  - Example: For $a = 18591$ and $b = 2291$, we have $18591 = 8 \cdot 2291 + 263$, so that $q = 8$ and $r = 263$.
  - The proof of the existence of q and r relies on the well-ordering principle, and can be shown using induction. Uniqueness follows by rearranging $qb + r = a = q'b + r'$ to obtain $r - r' = b(q' - q)$: since $-b < r - r' < b$, this means $q' - q$ is an integer between -1 and 1, and hence must be 0.

- Definition: If $d|a$ and $d|b$, then d is a common divisor of a and b. If a and b are not both zero, then there are only a finite number of common divisors: the largest one is called the greatest common divisor, or gcd, and denoted by $\gcd(a, b)$.

  - Warning: Many authors use the notation (a, b) to denote the gcd of a and b: this stems from the notation used for ideals in ring theory. The author of these notes generally dislikes using this notation and will write gcd explicitly, since otherwise it is easy to confuse the gcd with an ordered pair (a, b).
  - Example: The positive divisors of 30 are 1, 2, 3, 5, 6, 10, 15, 30. The positive divisors of 42 are 1, 2, 3, 6, 7, 14, 21, 42. The common (positive) divisors are 1, 2, 3, and 6, and the gcd is therefore 6.

- Definition: If $\gcd(a, b) = 1$, we say a and b are relatively prime. For example, 5 and 12 are relatively prime.

- Here are a few useful facts about greatest common divisors:

  - If $m > 0$, then $m \cdot \gcd(a, b) = \gcd(ma, mb)$.
  - If $d > 0$ divides both a and b, then $\gcd(a/d, b/d) = \gcd(a, b)/d$.
  - If a and b are both relatively prime to m, then so is ab.
  - For any integer x, $\gcd(a, b) = \gcd(a, b + ax)$.
  - If $c|ab$ and b, c are relatively prime, then $c|a$.

- It may seem difficult to compute the gcd of two large integers, since the most natural procedure would be to write down lists of all divisors of the two integers and then compare them. However, there is a much more efficient method for computing gcds:

- Theorem (Euclidean Algorithm): Given integers $0 < b < a$, repeatedly apply the division algorithm as follows, until a remainder of zero is obtained:

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

  Then $\gcd(a, b)$ is equal to the last nonzero remainder, r_{k+1}.

  - Proof: First observe that the algorithm will eventually terminate, because $b > r_1 > r_2 > \cdots \geq 0$, and the well-ordering principle dictates that there cannot exist an infinite decreasing sequence of nonnegative integers.
  - We now claim that $\gcd(a, b) = \gcd(b, r_1)$: this follows because $\gcd(b, r_1) = \gcd(b, a - q_1 b) = \gcd(b, a)$ from the gcd properties above.

- Now we can repeatedly apply this fact to see that $\gcd(a,b) = \gcd(b,r_1) = \gcd(r_1,r_2) = \cdots = \gcd(r_k,r_{k+1}) = r_{k+1}$ since r_{k+1} divides r_k.

- Corollary (GCD as a Linear Combination): If $d = \gcd(a,b)$, then there exist integers x and y with $d = xa+yb$.
  - Proof: By rearranging each equation in the Euclidean algorithm, we see that the newest remainder is a linear combination of the two previous terms.
  - By an easy induction, we therefore see that every remainder can be written as an explicit linear combination of a and b (since the first two remainders clearly can be so written). In particular, $r_{k+1} = xa+yb$ for some integers x and y.

- Example: Find the gcd of 1598 and 4879 using the Euclidean algorithm, and write the gcd explicitly as a linear combination of 1598 and 4879.
  - First, we use the Euclidean algorithm:

$$\begin{aligned} 4879 &= 3 \cdot 1598 + 85 \\ 1598 &= 18 \cdot 85 + 68 \\ 85 &= 1 \cdot 68 + 17 \\ 68 &= 4 \cdot 17 \end{aligned}$$

    and so the gcd is $\boxed{17}$.
  - For the linear combination, we solve for the remainders:

$$\begin{array}{rcccc} 85 &=& & = & 1 \cdot 4879 - 3 \cdot 1598 \\ 68 &=& 1598 - 18 \cdot 85 & = & -18 \cdot 4879 + 55 \cdot 1598 \\ 17 &=& 85 - 1 \cdot 68 & = & 19 \cdot 4879 - 58 \cdot 1598 \end{array}$$

    so we obtain $\boxed{17 = 19 \cdot 4879 - 58 \cdot 1598}$.

- The other fundamental fact about the integers is that they possess "unique prime factorization".

- Definition: If $p > 1$ is an integer, we say it is prime if there is no d with $1 < d < p$ such that $d|p$: in other words, if p has no positive divisors other than 1 and itself. If $n > 1$ is not prime, meaning that there is some $d|n$ with $1 < d < n$, we say n is composite. (The integer 1 is neither prime nor composite.)
  - The first few primes are 2, 3, 5, 7, 11, 13, 17, 19, and so forth.
  - Remark: In more advanced contexts, the following equivalent definition of a prime is often used instead: the integer $p > 1$ is prime if and only if $p|ab$ implies that $p|a$ or $p|b$.

- The prime numbers are often called the "building blocks under multiplication", because every positive integer can be written as the product of prime numbers in an essentially unique way:

- Theorem (Fundamental Theorem of Arithmetic): Every integer $n > 1$ can be factored into a product of primes, and this factorization is unique up to reordering of the factors.
  - To save space, we group equal primes together when actually writing out the canonical prime factorization: thus, $12 = 2^2 \cdot 3$, $720 = 2^2 \cdot 3^2 \cdot 5$, and so forth.
  - Proof: The existence of a prime factorization follows by induction on n: if $n > 1$ is prime, we are done, and if $n > 1$ is not prime, then n is the product of two smaller positive integers, which in turn have prime factorizations.
  - Uniqueness also follows inductively, by showing that any two factorizations of n must have at least one prime in common (cancelling it then allows an appeal to the inductive hypothesis).

- Prime factorizations yield an easy avenue for discussing divisibility and gcds:

- Proposition (Divisibility and Factorizations): If $a = \prod_{i=1}^{j} p_i^{a_i}$ and $b = \prod_{i=1}^{j} p_i^{b_i}$ for distinct primes p_i, then $a|b$ if and only if $a_i \leq b_i$ for each i. In particular, $\gcd(a,b) = \prod_{i=1}^{j} p_i^{\min(a_i,b_i)}$.

- Proof: We observe that if $b = ak$ and $k = \prod_{i=1}^{j} p_i^{k_i}$, then $a_i + k_i = b_i$. Since all exponents are nonnegative, saying that such an integer k exists is equivalent to saying that $a_i \leq b_i$ for all i.
- The statement about the gcd is immediate, since the exponent of p_i in the gcd is the largest integer that is $\leq a_i$ and $\leq b_i$, which is simply the minimum of a_i and b_i.

1.1.2 Modular Congruences and $\mathbb{Z}/m\mathbb{Z}$

- Definition: If m is a positive integer and m divides $b - a$, we say that a and b are congruent modulo m (or equivalent modulo m), and write "$a \equiv b \pmod{m}$".
  - Notation: As shorthand we usually write "$a \equiv b \pmod{m}$", or even just "$a \equiv b$" when the modulus m is clear from the context.
  - The statement $a \equiv b \pmod{m}$ can be thought of as saying "a and b are equal, up to a multiple of m".
  - Observe that if $m|(b-a)$, then $(-m)|(b-a)$ as well, so we do not lose anything by assuming that the modulus m is positive.
  - Example: $3 \equiv 9 \pmod{6}$, since 6 divides $9 - 3 = 6$.
  - Example: $-2 \equiv 28 \pmod{5}$, since 5 divides $28 - (-2) = 30$.
  - Example: $0 \equiv -666 \pmod{3}$, since 3 divides $-666 - 0 = -666$.
  - If m does not divide $b - a$, we say a and b are not congruent mod m, and write $a \not\equiv b \pmod{m}$.
  - Example: $2 \not\equiv 7 \pmod{3}$, because 3 does not divide $7 - 2 = 5$.
- Modular congruences share a number of properties with equalities:
- Proposition (Modular Congruences): For any positive integer m and any integers a, b, c, d, the following are true:
  1. $a \equiv a \pmod{m}$.
  2. $a \equiv b \pmod{m}$ if and only if $b \equiv a \pmod{m}$.
  3. If $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.
  4. If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$.
  5. If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $ac \equiv bd \pmod{m}$.
  6. If $a \equiv b \pmod{m}$, then $ac \equiv bc \pmod{mc}$ for any $c > 0$.
  7. If $d|m$, then $a \equiv b \pmod{m}$ implies $a \equiv b \pmod{d}$.
- The first three properties above demonstrate that modular equivalence is an equivalence relation[3].
- We would now like to study "arithmetic modulo m". To do this, we need to define the underlying objects of study:
- Definition: If a is an integer, the residue class of a modulo m, denoted $\overline{a}$, is the collection of all integers congruent to a modulo m. Observe that $\overline{a} = \{a + km,\ k \in \mathbb{Z}\}$.
- Example: The residue class of 2 modulo 4 is the set $\{\ldots, -6, -2, 2, 6, 10, 14, \ldots\}$, while the residue class of 2 modulo 5 is the set $\{\ldots, -8, -3, 2, 7, 12, 17, \ldots\}$.
- Here are a few fundamental properties of residue classes:
- Proposition (Properties of Residue Classes): Suppose m is a positive integer. Then

[3] A binary relation $\sim$ defined on a nonempty set S is called an equivalence relation if it obeys the following three axioms:

[E1] For any $a \in S$, $a \sim a$.

[E2] For any $a, b \in S$, $a \sim b$ implies $b \sim a$.

[E3] For any $a, b, c \in S$, $a \sim b$ and $b \sim c$ implies $a \sim c$.

Example: Equality of elements in any set (e.g., equality of real numbers) is an equivalence relation.

1. If a and b are integers with respective residue classes $\overline{a}$, $\overline{b}$ modulo m, then $a \equiv b \pmod{m}$ if and only if $\overline{a} = \overline{b}$.
   - Proof: If $\overline{a} = \overline{b}$, then by definition b is contained in the residue class $\overline{a}$, meaning that $b = a + km$ for some k. Thus, m divides $b - a$, so $a \equiv b \pmod{m}$.
   - Conversely, suppose $a \equiv b \pmod{m}$. If c is any element of the residue class $\overline{a}$, then by definition $c \equiv a \pmod{m}$, and therefore $c \equiv b \pmod{m}$.
   - Therefore, c is an element of the residue class $\overline{b}$, but since c was arbitrary, this means that $\overline{a}$ is contained in $\overline{b}$.
   - By the same argument with a and b interchanged, we see that $\overline{b}$ is also contained in $\overline{a}$, and thus $\overline{a} = \overline{b}$.
2. Two residue classes modulo m are either disjoint or identical.
   - Proof: Suppose that $\overline{a}$ and $\overline{b}$ are two residue classes modulo m. If they are disjoint, we are done, so suppose there is some c contained in both.
   - Then $c \equiv a \pmod{m}$ and $c \equiv b \pmod{m}$, so $a \equiv b \pmod{b}$. Then by property (1), we conclude $\overline{a} = \overline{b}$.
3. There are exactly m distinct residue classes modulo m, given by $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.
   - Proof: By the division algorithm, for any integer a there exists a unique r with $0 \leq r < m$ such that $a = qm + r$ with $q \in \mathbb{Z}$.
   - Then $a \equiv r \pmod{m}$, and so every integer is congruent modulo m to precisely one of the m integers $0, 1, \ldots, m-1$, which is to say, every integer lies in precisely one of the residue classes $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.

- Definition: The collection of residue classes modulo m is denoted $\mathbb{Z}/m\mathbb{Z}$ (read as "$\mathbb{Z}$ modulo $m\mathbb{Z}$").
  - Notation: Many other authors denote this collection of residue classes modulo m as $\mathbb{Z}_m$. We will avoid this notation and exclusively use $\mathbb{Z}/m\mathbb{Z}$ (or its shorthand $\mathbb{Z}/m$), since $\mathbb{Z}_m$ is used elsewhere in algebra and number theory for a different object.
  - By our properties above, $\mathbb{Z}/m\mathbb{Z}$ contains exactly m elements $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.
- We can now write down "addition and multiplication" modulo m using the residue classes of $\mathbb{Z}/m\mathbb{Z}$.
  - The fact that $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$ imply $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$ tell us that if we want to compute $a + c$ modulo m, then no matter which element b in the residue class of a and which element d in the residue class of c we take, the sum $b + d$ will lie in the same residue class as $a + c$, and the product bd will lie in the same residue class as ac.
  - Thus, everything makes perfectly good sense if we label the residue classes with the integers 0 through $m - 1$ and simply do the arithmetic with those residue classes.
- Definition: The addition operation in $\mathbb{Z}/m\mathbb{Z}$ is defined as $\overline{a} + \overline{b} = \overline{a + b}$, and the multiplication operation is defined as $\overline{a} \cdot \overline{b} = \overline{ab}$.
  - Here are the addition and multiplication tables for $\mathbb{Z}/5\mathbb{Z}$:

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{4}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	$\overline{3}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{1}$	$\overline{4}$	$\overline{2}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

  - Note that, for example, the statement $\overline{2} + \overline{4} = \overline{1}$ is now perfectly acceptable (and correctly stated with the equals sign): it says that if we take any element in the residue class $\overline{2}$ (modulo 5) and add it to any element in the residue class $\overline{4}$ (modulo 5), the result will always lie in the residue class $\overline{1}$ (modulo 5).

- Here are the addition and multiplication tables for $\mathbb{Z}/4\mathbb{Z}$:

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{0}$	$\overline{1}$	$\overline{2}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{0}$	$\overline{2}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

- Arithmetic modulo m is commonly described by ignoring residue classes entirely and only working with the integers 0 through $m-1$, with the result of every computation "reduced modulo m" to obtain a result lying in this range.
  - Thus, for example, to compute $3+10$ modulo 12, we would add to get 13 and then "reduce", yielding 1 modulo 12. Similarly, to find $3 \cdot 10$ modulo 12, we compute $3 \cdot 10 = 30$ and then reduce to obtain a result of 6 modulo 12.
  - However, this is a rather cumbersome and inelegant description. This definition is often used in programming languages, where "a mod m", frequently denoted "$a\%m$", is defined to be a *function* returning the corresponding remainder in the interval $[0, m-1]$.
  - Observe that with this definition, it is not true that $(a+b)\%m = (a\%m) + (b\%m)$, nor is is true that $ab\%m = (a\%m) \cdot (b\%m)$, since the sum and product may each exceed m. Instead, to obtain an actually true statement, one would have to write something like $ab\%m = [(a\%m) \cdot (b\%m)]\%m$.
  - In order to avoid such horrible kinds of statements, the best viewpoint really is to think of the statement $a \equiv b \pmod{m}$ as a congruence that is a "weakened" kind of equality, rather than always reducing each of the terms to its residue in the set $\{0, 1, \ldots, m-1\}$.
  - The other reason we adopt the use of residue classes is that they extend quite well to more general settings where we may not have such an obvious set of "representatives".
- The arithmetic in $\mathbb{Z}/m\mathbb{Z}$ shares many properties with the arithmetic in $\mathbb{Z}$ (which should not be surprising, since $\mathbb{Z}/m\mathbb{Z}$ was constructed using $\mathbb{Z}$):
- Proposition (Basic Arithmetic in $\mathbb{Z}/m\mathbb{Z}$): For any positive integer m the following properties of residue classes in $\mathbb{Z}/m\mathbb{Z}$ hold:
  1. The operation $+$ is associative: $\overline{a} + (\overline{b} + \overline{c}) = (\overline{a} + \overline{b}) + \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
  2. The operation $+$ is commutative: $\overline{a} + \overline{b} = \overline{b} + \overline{a}$ for any $\overline{a}$ and $\overline{b}$.
  3. The residue class $\overline{0}$ is an additive identity: $\overline{a} + \overline{0} = \overline{a}$ for any $\overline{a}$.
  4. Every residue class $\overline{a}$ has an additive inverse $-\overline{a}$ satisfying $\overline{a} + (-\overline{a}) = \overline{0}$.
  5. The operation $\cdot$ is associative: $\overline{a} \cdot (\overline{b} \cdot \overline{c}) = (\overline{a} \cdot \overline{b}) \cdot \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
  6. The operation $\cdot$ is commutative: $\overline{a} \cdot \overline{b} = \overline{b} \cdot \overline{a}$ for any $\overline{a}$ and $\overline{b}$.
  7. The operation $\cdot$ distributes over $+$: $\overline{a} \cdot (\overline{b} + \overline{c}) = \overline{a} \cdot \overline{b} + \overline{a} \cdot \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
  8. The residue class $\overline{1}$ is a multiplicative identity: $\overline{1} \cdot \overline{a} = \overline{a}$ for any $\overline{a}$.
     - Proof: For (1), by definition we have $\overline{a} + (\overline{b} + \overline{c}) = \overline{a} + \overline{b+c} = \overline{a + (b+c)}$ and also $(\overline{a} + \overline{b}) + \overline{c} = \overline{a+b} + \overline{c} = \overline{(a+b)+c}$.
     - But by the associative property [A1] in $\mathbb{Z}$, we know that $a + (b+c) = (a+b) + c$, so the associated residue classes are also equal.
     - The other properties follow in a similar way from the corresponding properties of the integers.
- The arithmetic in $\mathbb{Z}/m\mathbb{Z}$ shares many properties with the arithmetic in $\mathbb{Z}$. However, there are some very important differences.
  - For example, if a, b, c are integers with $ab = ac$ and $a \neq 0$, then we can "cancel" a from both sides to conclude that $b = c$.
  - However, this does not always work in $\mathbb{Z}/m\mathbb{Z}$: for example, $2 \cdot 1 = 2 \cdot 4$ modulo 6, but $1 \neq 4$ modulo 6.

- The issue here is that 2 and the modulus 6 are not relatively prime: 6 divides $2(4-1)$, but 6 does not divide $4-1$.

- We can explain the issue using modular congruences:

- Theorem (Invertible Elements in $\mathbb{Z}/m\mathbb{Z}$): If $m > 0$, then the residue class $\overline{a}$ has a multiplicative inverse in $\mathbb{Z}/m\mathbb{Z}$ if and only if a and m are relatively prime.

  - Proof: First suppose that a and m are relatively prime. Then by our analysis of the Euclidean algorithm, there exist integers x and y such that $xa + ym = 1$: then $xa \equiv 1 \pmod{m}$, which is to say $\overline{x} \cdot \overline{a} = \overline{1}$, so that $\overline{a}$ has a multiplicative inverse as claimed.
  - Conversely, suppose $\overline{a}$ were invertible in $\mathbb{Z}/m\mathbb{Z}$ with inverse $\overline{x}$. Then we would have $\overline{x} \cdot \overline{a} = \overline{1}$, or equivalently $xa \equiv 1 \pmod{m}$, and this is in turn equivalent to saying there exists an integer y with $xa + ym = 1$. But then the common divisor d would divide $xa + ym$ hence divide 1, and so a and m are relatively prime.

- The proof above shows that we can compute the inverse of an invertible residue class using the Euclidean algorithm.

- Example: Find the multiplicative inverse of $\overline{9}$ in $\mathbb{Z}/11\mathbb{Z}$.

  - Using the Euclidean algorithm, we can obtain $1 = 5 \cdot 11 - 6 \cdot 9$. Reducing both sides modulo 11 yields $\overline{1} = \overline{-6} \cdot \overline{9}$, and since $\overline{-6} = \overline{5}$, this shows that the multiplicative inverse of $\overline{9}$ in $\mathbb{Z}/11\mathbb{Z}$ is $\boxed{\overline{5}}$.

- The case where the modulus is prime is of particular importance:

- Corollary: If p is a prime number, then every nonzero residue class in $\mathbb{Z}/p\mathbb{Z}$ has a multiplicative inverse.

  - Proof: If p is prime, then p is relatively prime to each of 1, 2, ... , $p-1$ and hence all of the nonzero residue classes modulo p are invertible.

- Equivalently, this corollary states that $\mathbb{Z}/p\mathbb{Z}$ is a field.

  - We will discuss the structure of fields at great length later, but to summarize: a field is a set F together with two binary operations of addition ($+$) and multiplication ($\cdot$) both of which are associative and commutative and where $\cdot$ distributes over $+$, that also possesses an additive identity 0 and a multiplicative identity $1 \neq 0$, and where every element has an additive inverse and every nonzero element has a multiplicative inverse.
  - Standard (and likely familiar) examples of fields include the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$.
  - Frequently, the field $\mathbb{Z}/p\mathbb{Z}$ is also denoted $\mathbb{F}_p$ ("the field with p elements"), and sometimes in older literature as $GF(p)$ ("the Galois Field with p elements") since these fields were first extensively studied by Évariste Galois. The uses of the word "the" are justified, as we will show later in our study of fields that (up to relabeling the elements), there is only one field with p elements for any prime p.

1.2 Polynomials

- We begin by discussing the structure of polynomials whose coefficients lie in a field.

  - Polynomials with real coefficients (like $p(x) = 1 + x^2$ or $q(x) = 3 + \pi x^2$) are likely familiar from elementary algebra.
  - Unlike in elementary algebra, however, our polynomials will be "formal symbols" rather than functions. We will soon exploit the connection between polynomials and functions, but, as we will discuss, there are very important reasons for us to take a more abstract approach to polynomials than simply viewing them as functions.

- Definition: Let F be a field and x be an indeterminate. A polynomial in x with coefficients in F consists of a formal sum $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, for an integer $n \geq 0$ and where each element $a_i \in F$.

- The term "indeterminate" is deliberately undefined in the definition above. A more concrete[4] (but vastly less intuitive) definition of polynomials can be given using Cartesian products, but we will not use it.
- If $a_n \neq 0$, we say that the polynomial has degree n and if $a_n = 1$ we say the polynomial is monic. (By convention, the degree of the zero polynomial 0 is $-\infty$.)
- The leading term of the polynomial is its highest-degree term (i.e., $a_n x^n$) and its leading coefficient is the corresponding coefficient (i.e., a_n).
- We will employ the traditional "function" notation for polynomials (e.g., by writing a polynomial as $p(x) = x^2 + 5$), and also often drop the variable portion (e.g., by referring to "the polynomial p") when convenient. We reiterate, however, that our polynomials are *not* functions, but rather formal sums.

- As is familiar from elementary algebra, the polynomials with coefficients in F have natural arithmetic operations:

- Definition: The set of polynomials in x with coefficients in F, denoted $F[x]$, has the two operations of addition and multiplication defined as follows:
  - Addition is defined "termwise":

$$(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0) + (b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0) = (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \cdots + (a_0 + b_0).$$

  - Multiplication is defined first on "monomials" (polynomials with only one nonzero coefficient), via $(ax^n) \cdot (bx^m) = abx^{n+m}$, and then extended to arbitrary polynomials via the distributive laws. Explicitly,

$$(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n) \cdot (b_0 + b_1 x + b_2 x^2 + \cdots + b_m x^m) = a_0 b_0 + (a_1 b_0 + a_0 b_1)x + (a_2 b_0 + a_1 b_1 + a_0 b_2)x^2 + \cdots + a_n b_m x^{n+m}$$

    where the coefficient of x^j in the product is given by $\sum_{k=0}^{j} a_k b_{j-k}$.
  - It is tedious (but not difficult) to verify the basic properties of arithmetic for $F[x]$: for example, that polynomial addition and multiplication are commutative and associative, that multiplication distributes over addition, that the polynomial 0 is an additive identity and the polynomial 1 is a multiplicative identity, and so forth.

- Example: In $\mathbb{Q}[x]$, find $p(x) + q(x)$ and $p(x) \cdot q(x)$ for $p(x) = 2 + 3x$ and $q(x) = 3 + 2x$.
  - We have $p(x) + q(x) = \boxed{5 + 5x}$, while $p(x)q(x) = 6 + (4 + 9)x + 6x^2 = \boxed{6 + 13x + 6x^2}$.

- Example: In $\mathbb{F}_5[x]$, find $p(x) + q(x)$ and $p(x) \cdot q(x)$ for $p(x) = 2 + 3x$ and $q(x) = 3 + 2x$.
  - We have $p(x) + q(x) = 5 + 5x = \boxed{0}$, while $p(x)q(x) = 6 + (4 + 9)x + 6x^2 = \boxed{1 + 3x + x^2}$.

- Degrees behave quite well under addition and multiplication:

- Proposition (Properties of Degree): If p and q are any polynomials in $F[x]$, then $\deg(p+q) \leq \max(\deg p, \deg q)$, and $\deg(p \cdot q) = \deg p + \deg q$.
  - Proof: It is straightforward to verify that each claim holds if p or q is zero (in which case the left side of each inequality is $-\infty$). Now assume p and q are nonzero.
  - For $p + q$, observe that if there are no terms of degree n or higher in p or q, then there are no terms of degree n or higher in $p + q$ either.
  - For $p \cdot q$, observe that if the leading terms of p and q are $a_n x^n$ and $b_m x^m$ respectively, then the leading term of $p \cdot q$ is $a_n b_m x^{m+n}$, and $a_n b_m \neq 0$ since F is a field.

[4] Specifically: inside the Cartesian product $\prod_{\mathbb{Z}_{\geq 0}} F = (a_0, a_1, a_2, \dots)$ indexed by the nonnegative integers, we define the "polynomials" to be the sequences all but finitely many of whose entries are zero, and interpret the sequence $(a_0, a_1, a_2, \dots, a_n, 0, 0, \dots)$ as the formal sum $a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$. We can then define the operations of polynomial addition and multiplication solely in terms of these sequences.

1.2.1 The Division Algorithm and Euclidean Algorithm in $F[x]$

- We can define divisibility of polynomials in the same way as in $\mathbb{Z}$:
- Definition: If $a, b \in F[x]$, we say that a divides b (written $a|b$), if there is a $k \in F[x]$ with $b = ka$.
    - Example: We see that $x-1$ divides x^2-1 in $\mathbb{Q}[x]$, since $x^2-1 = (x-1)(x+1)$.
- Much like $\mathbb{Z}$, $F[x]$ also possesses a "long division" algorithm: the only difference is that we measure the "size" of a polynomial via its degree.
- Theorem (Division Algorithm in $F[x]$): If F is any field, and $a(x)$ and $b(x)$ are any polynomials in $F[x]$ with $b(x) \neq 0$, then there exist unique polynomials $q(x)$ and $r(x)$ such that $a(x) = b(x)q(x) + r(x)$, where $\deg(r) < \deg(b)$. Furthermore, $b|a$ if and only if $r = 0$.
    - 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 in order 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.)
    - For example, when we divide the polynomial x^3+x^2+3x+5 by the polynomial x^2+3x+1 in $\mathbb{R}[x]$, we obtain the quotient $q(x) = x-2$ and remainder $r(x) = 8x+7$: indeed, we have $x^3+x^2+3x+5 = (x-2)(x^2+3x+1)+(8x+7)$.
    - 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 $a(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ and $b(x) = b_m x^m + \cdots + b_0$, 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.
    - For the uniqueness, suppose that $a = qb + r = q'b + r'$: then $r - r' = b(q'-q)$ has degree less than $\deg(b)$ but is also divisible by b, hence must be zero.
    - Finally, by definition if $r = 0$ then $b|a$, and conversely if $b|a$ then since r is unique we must have $r = 0$.
- The existence of this division algorithm in $F[x]$ allows us to adapt many results that hold in $\mathbb{Z}$ into this setting. First is the idea of a common divisor:
- Definition: If a and b are polynomials in $F[x]$, we say a polynomial d is a common divisor if $d|a$ and $d|b$.
    - Example: The polynomials $x+1$ and $2x+2$ are both common divisors of x^2-1 and x^2+3x+2 in $\mathbb{R}[x]$.
- We would naturally want to define the greatest common divisor to be the polynomial of largest degree dividing both a and b.
    - However, this polynomial is not unique: in the example above, it is easy to see that x^2-1 and x^2+3x+2 do not have a common divisor of degree 2 (or larger), so both $x+1$ and $2x+2$ are common divisors of maximal degree.
    - Ultimately, $x+1$ and $2x+2$ are essentially the same (as far as divisibility goes), since they only differ by a constant factor. This situation occurs often enough that we give it a name:
    - Definition: If p and q are polynomials in $F[x]$ and there exists a nonzero constant c such that $p = cq$, we say p and q are associate.

- Definition: If a and b are polynomials in $F[x]$, not both zero, we say the polynomial d is a greatest common divisor of a and b if it a common divisor of a and b with the property that if d' is any other common divisor, then $d'|d$.
  - Under this definition, we can verify that both $x+1$ and $2x+2$ are gcds of x^2-1 and x^2+3x+2.
- This definition does not immediately imply that a gcd actually exists. But by adapting the Euclidean algorithm to this setting, we can give a procedure for computing the gcd (and in particular, implying that it exists and is essentially unique) and for writing it as a linear combination:
- Algorithm (Euclidean Algorithm in $F[x]$): Given polynomials a and b in $F[x]$, not both zero, repeatedly apply the division algorithm as follows, until a remainder of zero is obtained:

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

  Then the last nonzero remainder r_{k+1} is a gcd of a and b. Furthermore, by successively solving for the remainders and plugging in the previous equations, r_{k+1} can be explicitly written as a linear combination of a and b.
  - Proof: First observe that the algorithm will eventually terminate with a zero remainder, because $\deg(b) > \deg(r_1) > \deg(r_2) > \cdots$, and the well-ordering principle dictates that there cannot exist an infinite decreasing sequence of nonnegative integers.
  - By an easy induction, we can see that if $d|a$ and $d|b$, then $d|r_j$ for each $j \geq 1$: thus, any common divisor of a and b must divide r_{k+1}.
  - Conversely, by another easy induction, r_{k+1} divides each r_j for each $j \geq 1$, and thus r_{k+1} divides both a and b.
  - Therefore, r_{k+1} divides both a and b, and any other common divisor also divides r_{k+1}: thus, r_{k+1} is a gcd of a and b.
  - The correctness of the algorithm for computing the gcd as a linear combination follows by an easy induction.
- If a and b are not both zero, we can make the gcd unique by additionally requiring that it be monic (i.e., have leading coefficient 1).
  - Explicitly, if d_1 and d_2 are both greatest common divisors of a and b, then $d_1|d_2$ and $d_2|d_1$, so that $d_1 = sd_2$ and $d_2 = td_1$ for some polynomials s and t.
  - By comparing degrees, we see that $\deg(s) = \deg(t) = 0$, meaning that s and t must both be constants, and thus d_1 and d_2 are associates. In particular, there is a unique gcd whose leading coefficient is 1.
- Example: Find "the" 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 "the" 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 $1 = \boxed{(-\frac{1}{2} - \frac{3}{20}) \cdot (x^3 + 7x^2 + 9x - 2) + (\frac{21}{20} + \frac{19}{20}x + \frac{3}{20}x^2) \cdot (x^2 + 4x)}$.

1.2.2 Irreducible Polynomials and Unique Factorization

- We next develop the polynomial analogue of the prime factorization of an integer: namely, writing a polynomial as a product of irreducible factors, and showing that this factorization is essentially unique.
- Definition: A nonzero polynomial $p \in F[x]$ is irreducible if it is not a constant, and for any "factorization" $p = bc$ with $b, c \in F[x]$, one of b and c must be a constant polynomial. If p is not a constant and possesses a factorization $p = bc$ where neither b nor c is constant, then p is reducible.
  - Equivalently, a polynomial is irreducible if it cannot be written as a product of two polynomials of smaller positive degree, and is reducible if it can be so written.
  - Example: Any polynomial of degree 1 is irreducible.
  - 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 in $F[x]$ depends on the field F. 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 $\mathbb{C}[x]$.
- The irreducible polynomials are the "building blocks under multiplication" in $F[x]$, much like the primes are in $\mathbb{Z}$, because every nonzero polynomial can be written as the product of irreducibles:
- Proposition (Factorization into Irreducibles): Every polynomial of positive degree in $F[x]$ can be written as a product of irreducible polynomials (where a "product" is allowed to have only one term).
  - Proof: We use strong induction on $n = \deg(p)$. The result clearly holds if $n = 1$, since any polynomial of degree 1 is irreducible.
  - Now suppose $n \geq 2$. If p is irreducible, we are done, so otherwise assume that p is reducible.
  - By definition, there exist polynomials a, b with $0 < \deg(a), \deg(b) < n$ with $p = ab$.

- By the strong induction hypothesis, both a and b can be written as a product of irreducibles; multiplying these two products then gives p as a product of irreducibles.

- In order to show that the factorization into irreducibles is unique, we need the analogous divisibility property that we required in $\mathbb{Z}$:

- Proposition (Irreducibles are Prime in $F[x]$): If $p \in F[x]$ is irreducible and $p|ab$, then $p|a$ or $p|b$.

  - Proof: Suppose $p|ab$. If $p|a$, we are done, so suppose $p \nmid a$, and let d be a gcd of p and a.
  - By hypothesis, d divides p, so (since p is irreducible) either d is a constant, or $d = up$ for some constant u: however, the latter cannot happen, because then up (hence p) would divide a.
  - Hence d is a constant, 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 $p|b$.

- Now we can prove that the factorization into irreducibles is essentially unique up to reordering.

  - There is one additional wrinkle that we must address, however, which we illustrate with an example.
  - In $\mathbb{C}[x]$, we can write $x^2 + 1 = (x + i)(x - i) = (ix + 1)(-ix + 1)$.
  - 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: $x + i = i(-ix + 1)$ and $x - i = (-i)(ix + 1)$.
  - We should declare that two factorizations are equivalent if the only differences between them are by reordering terms or moving constant factors around, which is equivalent to replacing elements with associates.

- Theorem (Unique Factorization in $F[x]$): Every polynomial of positive degree in $F[x]$ can be written as a product of irreducible polynomials. Furthermore, this factorization is unique up to reordering and associates: if $p = r_1 r_2 \cdots r_d = q_1 q_2 \cdots q_k$, then $d = k$ and there is some reordering of the factors such that p_i and q_i are associate for each $1 \leq i \leq k$.

  - Proof: We proved the existence of a factorization above. For the uniqueness, we induct on the number of irreducible factors of $p = r_1 r_2 \cdots r_d$.
  - If $d = 0$, then p is a constant. If p had some other factorization $p = rc$ with r irreducible, then q would divide a constant, hence be a constant (impossible).
  - Now suppose $d \geq 1$ and that $r = r_1 r_2 \cdots r_k = q_1 q_2 \cdots q_d$ has two factorizations into irreducibles.
  - Since $r_1|(q_1 \cdots q_d)$ and r_1 is irreducible, repeatedly applying the fact that r_1 irreducible and $r_1|ab$ implies $r_1|a$ or $r_1|b$ shows that r_1 must divide q_i for some i.
  - Then $q_i = r_1 u$ for some u: then since q_i is irreducible (and r_1 is not a constant), u must be a constant, and thus q_i and r_1 are associates.
  - Cancelling r_1 from both sides then yields the equation $r_2 \cdots r_d = (uq_2) \cdots q_k$, which is a product of fewer irreducibles. By the induction hypothesis, such a factorization is unique up to associates. This immediately yields the desired uniqueness result for p as well.

1.2.3 Roots of Polynomials, Irreducibility

- In elementary algebra, polynomials are examples of functions. We would like to extend this idea of "plugging values in" to a general polynomial in $F[x]$.

- 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 $a_0 + a_1 r + \cdots + a_n r^n \in F$.

  - 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 $p : F \to F$, with $p(r) = a_0 + a_1 r + \cdots + a_n r^n$.
- 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 $p = x + x^2$ 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 $(x^k - r^k) = (x - r)(x^{k-1} + x^{k-2} r + \cdots + x r^{k-2} + r^{k-1})$, 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 zeroes that a polynomial can have:

- Proposition: 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).

- In general, it is not easy to determine when an arbitrary polynomial is irreducible. If the degree is small, however, this task can be performed by examining all possible factorizations. The following result is frequently useful:

- Proposition (Irreducibility in Degrees 2 and 3): If F is a field and $p \in F[x]$ has degree 2 or 3 and has no zeroes in F, then p is irreducible.
  - Proof: If $p(x) = a(x)b(x)$, taking degrees shows $\deg(p) = \deg(a) + \deg(b)$. Since a and b both have positive degree and $\deg(p)$ is 2 or 3, at least one of a and b must have degree 1. Then its root is also a root of $p(x)$. Taking the contrapositive gives the desired statement.
  - Example: Over $\mathbb{R}$, the polynomial $x^2 + 2x + 11$ has no roots (it is always positive, as can be seen by completing the square), so it is irreducible.
  - Example: Over $\mathbb{F}_2$, the polynomial $q(x) = x^3 + x + 1$ is irreducible: it has no roots since $q(0) = q(1) = 1$.

- Example: Over $\mathbb{F}_5$, the polynomial $q(x) = x^3 + x + 1$ is irreducible: it has no roots since $q(0) = 1$, $q(1) = 3$, $q(2) = 1$, $q(3) = 1$, and $q(4) = 4$.
- Note of course that a polynomial of larger degree can be reducible without having any zeroes: for example, $x^4 + 3x^2 + 2$ has no zeroes in $\mathbb{R}$, but it is still reducible: $x^4 + 3x^2 + 2 = (x^2 + 1)(x^2 + 2)$.

- 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. (We will later be able to give a proof using other techniques from field and group theory.)
  - 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.

1.2.4 Factorization and Irreducibility in $\mathbb{Q}[x]$

- It is more difficult to test whether a polynomial is irreducible in $\mathbb{Q}[x]$. A central idea is that we can reduce the problem of factoring in $\mathbb{Q}[x]$ to one of factoring in $\mathbb{Z}[x]$, the set of polynomials with integer coefficients, by "clearing denominators".
  - Specifically, if p is any polynomial in $\mathbb{Q}[x]$, we may multiply p by the product of all the denominators of its coefficients (or their least common multiple) to obtain a polynomial in $\mathbb{Z}[x]$. Since every nonzero integer is invertible in $\mathbb{Q}$, the factorization of this new polynomial, with integer coefficients, will be essentially the same as that of the original polynomial.
  - As an example, consider the problem of factoring $p(x) = 2x^3 + x^2 + \frac{2}{3}x + \frac{1}{3}$ in $\mathbb{Q}[x]$.
  - Since 3 is an invertible constant in $\mathbb{Q}[x]$, we may equivalently ask about the factorization of $3p(x) = 6x^3 + 3x^2 + 2x + 1$ in $\mathbb{Z}[x]$.

- We start by proving the famous "rational root test", which allows us to determine whether a given polynomial in $\mathbb{Z}[x]$ has a rational root:
- Proposition (Rational Root Test): Suppose $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ is a polynomial in $\mathbb{Z}[x]$. Then any root r/s (in lowest terms) must have $r|a_0$ and $s|a_n$.
  - Proof: If r/s is a root of $p(x)$, then $a_n(r/s)^n + a_{n-1}(r/s)^{n-1} + \cdots + a_0 = 0$. Clearing denominators yields $a_n r^n + a_{n-1}r^{n-1}s + \cdots + a_1 rs^{n-1} + a_0 s^n = 0$.
  - Thus, by rearranging, we see that $a_n r^n = s(-a_{n-1}r^{n-1} - \cdots - a_0 s^{n-1})$, so s divides $a_n r^n$. But since s and r are relatively prime, this means s divides a_n.
  - In a similar way, since $a_0 s^n = r(-a_n r^{n-1} - \cdots - a_1 s^{n-1})$, we see that r divides $a_0 s^n$ hence a_0.

- This test allows us to make a finite list of possible rational roots for any polynomial with integer coefficients.
- Example: Show that the polynomial $p(x) = x^3 + ax + 1$ is irreducible in $\mathbb{Q}[x]$ for any integer $a \neq 0, -2$.
  - Since this polynomial has degree 3, we need only show that it has no roots in $\mathbb{Q}$.
  - By the rational root test, the only possible rational roots are ± 1, and since $p(1) = 2 + a$ and $p(-1) = a$, the conditions on a imply that p has no rational roots. Thus, p is irreducible.

- It is natural to think that factorization of polynomials in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$ are essentially "the same", and this is ultimately true (though not quite so easy to prove rigorously):
- Theorem (Gauss's Lemma): If $p(x) \in \mathbb{Z}[x]$ has positive degree and is reducible in $\mathbb{Q}[x]$, then $p(x) = f(x)g(x)$ for some $f(x), g(x) \in \mathbb{Z}[x]$ of positive degree.
  - Proof: We say a polynomial in $\mathbb{Z}[x]$ is "primitive" if the gcd of its coefficients is equal to 1.
  - First, we observe that, in $\mathbb{Q}[x]$, any nonzero polynomial $a(x)$ is associate to a primitive polynomial in $\mathbb{Z}[x]$.
  - To see this, let d be the least common multiple of the denominators of $a(x)$: then $d \cdot a(x)$ is a polynomial in $\mathbb{Z}[x]$. Now let e be the greatest common divisor of the coefficients of $d \cdot a(x)$: then $\frac{d}{e} \cdot a(x)$ is a primitive polynomial in $\mathbb{Z}[x]$; since $\frac{d}{e}$ is invertible in $\mathbb{Q}$, this primitive polynomial is associate to $a(x)$.
  - Next, we claim that the product of two primitive polynomials is also primitive.
  - To see this, suppose that $a(x)b(x)$ is not primitive for some $a(x), b(x) \in \mathbb{Z}[x]$, with $a(x) = a_0 + a_1 x + \cdots + a_n x^n$ and $b(x) = b_0 + \cdots + b_m x^m$: then since $a(x)b(x)$ is not primitive, all of its coefficients are divisible by some prime s.
  - If there is at least one coefficient of each of $a(x)$ and $b(x)$ not divisible by s, suppose that a_i and b_j are the lowest-degree such coefficients. Then the degree-$(i+j)$ term of $a(x)b(x)$ is $a_0 b_{i+j} + \cdots + a_{i-1} b_{j+1} + a_i b_j + a_{i+1} b_{j-1} + \cdots + a_{i+j} b_0$, but by hypothesis each term except $a_i b_j$ is divisible by s. This is a contradiction, since this coefficient of $a(x)b(x)$ would then not be divisible by s.
  - Now, returning to the original problem, suppose that $p(x)$ is reducible in $\mathbb{Q}[x]$, say as $p(x) = f_0(x)g_0(x)$ with f_0 and g_0 both of positive degree.
  - By our first observation, both f_0 and g_0 are associate to a primitive polynomial: say, f, and g respectively.
  - Then (by rearranging the corresponding unit factors) we see that $d \cdot p(x) = e \cdot f(x) \cdot g(x)$ for some relatively prime integers d and e.
  - Now notice that since d and e are relatively prime, d must divide all coefficients of $f(x)g(x)$. But $f(x)g(x)$ is primitive by our second observation, so we must have $d = \pm 1$.
  - Then $p(x) = [ed^{-1} \cdot f(x)] \cdot g(x)$ is a nontrivial factorization of $p(x)$ over $\mathbb{Z}[x]$, as required.
- As we can see from the proof above, roughly speaking (i.e., up to shuffling around constant factors), factoring in $\mathbb{Q}[x]$ is the same as factoring in $\mathbb{Z}[x]$.
  - The advantage to working in $\mathbb{Z}[x]$, however, is that we can exploit properties of $\mathbb{Z}$ to establish that factorizations cannot exist.
- Example: Show that the polynomial $p(x) = x^4 + x^3 - 2x^2 + x + 1$ is irreducible in $\mathbb{Q}[x]$.
  - First, by the rational root test, the only possible roots of this polynomial are ± 1, neither of which is a root.
  - Thus, if $p(x)$ were reducible, it would factor as a product of two quadratics. By moving factors of -1 around (as needed) such a factorization would have the form $p(x) = (x^2 + ax + b)(x^2 + cx + d)$.
  - By expanding and comparing coefficients, we see that $a + c = 1$, $b + ac + d = -2$, $ad + bc = 1$, and $bd = 1$.
  - The last equation gives $(b, d) = (1, 1)$ or $(-1, -1)$.
  - If $b = d = 1$ then we obtain the equations $a + c = 1$ and $ac = -4$, which has no integer solutions.
  - If $b = d = -1$ then we obtain $a + c = 1$, $ac = 0$, and $a + c = -1$, which has no solutions at all (integer or otherwise).
  - Therefore, $p(x)$ is irreducible, as claimed.
- Note, however, that a similar sort of analysis becomes very difficult in larger degree (and also becomes more difficult when the coefficients are large).

- For example, to show in this manner that a polynomial of degree 7 is irreducible, one would need to verify that it has no roots, as well as no factorization into a product of polynomials of degree 2 and 5, or 3 and 4.

- For this reason, other irreducibility criteria have been developed. Here is one:

- Theorem (Eisenstein-Schönemann Criterion): Let $q(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial in $\mathbb{Z}[x]$. If each coefficient $a_0, a_1, \ldots, a_{n-1}$ is divisible by a prime p, and a_0 is not divisible by p^2, then $q(x)$ is irreducible in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

  - Proof: Suppose that $q(x) = b(x)c(x)$ were reducible in $\mathbb{Z}[x]$, with $b(x) = x^s + b_{s-1}x^{s-1} + \cdots + b_0$ and $c(x) = x^t + c_{t-1}x^{t-1} + \cdots + c_0$.
  - Since p divides $a_0 = b_0 c_0$, p divides at least one of these coefficients: without loss of generality, suppose $p|b_0$.
  - Now let b_i be the lowest-degree coefficient of $b(x)$ not divisible by p (there must be one, since $b_s = 1$ is not divisible by p): then we have $a_i = b_0 c_i + b_1 c_{i-1} + \cdots + b_{i-1}c_1 + b_i c_0$.
  - Since p divides a_i and also divides the terms $b_0 c_i$, $b_1 c_{i-1}$, ... , $b_{i-1}c_1$, it must divide $b_i c_0$. But since p does not divide b_i, we see that p divides c_0.
  - But then p divides both b_0 and c_0, meaning that p^2 divides $b_0 c_0 = a_0$. This is a contradiction, so there cannot exist any such factorization of $q(x)$.
  - Thus, $q(x)$ is irreducible in $\mathbb{Z}[x]$, and then by Gauss's lemma, q is irreducible in $\mathbb{Q}[x]$ as well.

- Example: By Eisenstein's criterion with $p = 2$, the polynomial $x^n - 2$ is irreducible in $\mathbb{Z}[x]$ for any positive integer n.

- Example: Show that the polynomial $q(x) = x^4 + x^3 - 3x^2 + x + 7$ is irreducible in $\mathbb{Q}[x]$.

  - We cannot apply Eisenstein's criterion to this polynomial directly.
  - However, notice that $q(x-1) = x^4 - 3x^3 + 6x + 3$, and this polynomial is irreducible by Eisenstein's criterion with $p = 3$.
  - It is then easy to see that any factorization of $q(x-1)$ would give a factorization of $q(x)$, and vice versa: therefore, the original polynomial $q(x)$ must also have been irreducible.

- We can also use calculations in $\mathbb{F}_p[x]$ to show that a polynomial is irreducible in $\mathbb{Z}[x]$.

  - Specifically, if a polynomial factors in $\mathbb{Z}[x]$, then reducing the factorization modulo p yields a factorization in $\mathbb{F}_p[x]$, as long as the degrees of the factors do not change, which will be the case whenever the leading coefficient of the polynomial is not divisible by p.
  - By taking the contrapositive of the observation above, we see that if $q(x)$ is irreducible in $\mathbb{F}_p[x]$ and has leading coefficient not divisible by p, then it must also be irreducible in $\mathbb{Z}[x]$ (and thus by Gauss's lemma, also in $\mathbb{Q}[x]$).

- Example: Show that $q(x) = x^3 + 12x^2 + 27x + 345$ is irreducible in $\mathbb{Z}[x]$.

  - Notice that $q(x) \equiv x^3 + x + 1$ modulo 2, and so q has no roots modulo 2. Since q has degree 3, this means q is irreducible in $\mathbb{F}_2[x]$, and hence also in $\mathbb{Z}[x]$.

1.2.5 Polynomial Modular Arithmetic

- We now turn our attention to discussing modular congruences (and modular arithmetic) in $F[x]$.

- Our underlying definition of modular congruences and residue classes are exactly the same as over $\mathbb{Z}$:

- Definition: Let F be a field. If $a, b, p \in F[x]$, we say that a is congruent to b modulo p, written $a \equiv b \pmod{p}$, if $p|(b-a)$.

  - Example: In $\mathbb{R}[x]$, it is true that $x^2 \equiv x$ modulo $x - 1$, because $x - 1$ divides $x^2 - x = x(x-1)$.

- Example: In $\mathbb{F}_2[x]$, it is true that $x^3 + x \equiv x + 1$ modulo $x^2 + x + 1$, because $(x^2 + x + 1)$ divides $(x^3 + x) - (x + 1) = (x + 1)(x^2 + x + 1)$.

- Most of the basic properties of modular congruences in $\mathbb{Z}$ extend to $F[x]$ with little or no change:
- Proposition (Modular Congruences): Let F be a field. If $a, b, c, d, p \in F[x]$ and $p \neq 0$, then the following are true:

1. $a \equiv a \pmod{p}$.
2. $a \equiv b \pmod{p}$ if and only if $b \equiv a \pmod{p}$.
3. If $a \equiv b \pmod{p}$ and $b \equiv c \pmod{p}$, then $a \equiv c \pmod{p}$.
4. If $a \equiv b \pmod{p}$ and $c \equiv d \pmod{p}$, then $a + c \equiv b + d \pmod{p}$.
5. If $a \equiv b \pmod{p}$ and $c \equiv d \pmod{p}$, then $ac \equiv bd \pmod{p}$.
    - Each of these is a straightforward calculation using the definition of congruence.

- We can now construct residue classes, again in exactly the same way:
- Definition: If $a, r \in F[x]$, the residue class of a modulo r, denoted $\overline{a}$, is the set $S = \{a + dr : d \in F[x]\}$ of all elements in $F[x]$ congruent to a modulo r.
    - Example: The residue class of 1 modulo x in $\mathbb{F}_2[x]$ is $\{1, 1 + x, 1 + x^2, 1 + x + x^2, 1 + x^3, \dots\}$.
- Here are a few fundamental properties of residue classes:
- Proposition (Properties of Residue Classes): Let F be a field and suppose $p \in F[x]$ is nonzero. Then

1. If a and b are polynomials in $F[x]$, then $a \equiv b \pmod{p}$ if and only if $\overline{a} = \overline{b}$.
    - Proof: Identical to the proof over $\mathbb{Z}$.
2. Two residue classes modulo p are either disjoint or identical.
    - Proof: Identical to the proof over $\mathbb{Z}$.
3. The residue classes modulo p are precisely those of the form $\overline{r}$ where $\deg(r) < \deg(p)$.
    - Proof: By the division algorithm, for any polynomial a there exists a unique r with $\deg(r) < \deg(p)$ such that $a = qm + r$ with $q \in F[x]$.
    - Then $a \equiv r \pmod{p}$, and so every polynomial is congruent modulo p to precisely one polynomial r with $\deg(r) < \deg(p)$.
    - In other words, every polynomial is contained in exactly one of the residue classes $\overline{r}$ where $\deg(r) < \deg(p)$.
    - By property (2), we conclude that these are all the residue classes, and that they are disjoint.

- If F is an infinite field, then if $\deg(p) > 0$, there will always be infinitely many residue classes in $F[x]$ modulo $p(x)$.
    - However, when F is a finite field of cardinality $|F|$, then the residue classes are each represented by a unique polynomial in $F[x]$ of degree less than $\deg(p)$.
    - Such a polynomial has exactly $\deg(p)$ coefficients (for the terms of degree 0, 1, ... , $\deg(p) - 1$), and each coefficient has $|F|$ possible choices: thus, there are precisely $|F|^{\deg(p)}$ residue classes modulo $p(x)$.

- Example: List the residue classes in $\mathbb{F}_2[x]$ modulo x^2.
    - Each coefficient is either 0 or 1, and by the above result, the residue classes are precisely the polynomials of degree less than 2.
    - Thus, there are four residue classes in $\mathbb{F}_2[x]$ modulo x^2: $\overline{0}$, $\overline{1}$, $\overline{x}$, and $\overline{x+1}$.

- Definition: If F is a field and $p \in F[x]$ is nonzero, the set of residue classes modulo p is denoted as $F[x]/p$ (read as "$F[x]$ modulo p").

- Like in $\mathbb{Z}/m\mathbb{Z}$, we have natural addition and multiplication operations in $F[x]/p$:
- Definition: The addition operation in $\mathbb{Z}/m\mathbb{Z}$ is defined as $\overline{a} + \overline{b} = \overline{a+b}$, and the multiplication operation is defined as $\overline{a} \cdot \overline{b} = \overline{ab}$.
    - In order for this definition to make sense, we need to verify that these operations are well-defined: that is, if we choose different elements $a' \in \bar{a}$ and $b' \in \bar{b}$, the residue class of $a' + b'$ is the same as that of $a + b$, and similarly for the product.
    - To see this, if $a' \in \bar{a}$ then $a' \equiv a \pmod{p}$, and similarly if $b' \in \bar{b}$ then $b' \equiv b \pmod{p}$.
    - Then $a' + b' \equiv a + b \pmod{p}$, so $\overline{a'+b'} = \overline{a+b}$. Likewise, $a'b' \equiv ab \pmod{p}$, so $\overline{a'b'} = \overline{ab}$.
    - Thus, the operations are well-defined.
- Proposition (Basic Arithmetic in $F[x]/p$): Let F be a field and $p \in F[x]$ be nonzero. Then the following properties hold for residue classes in $F[x]/p$:
    1. The operation $+$ is associative: $\overline{a} + (\overline{b} + \overline{c}) = (\overline{a} + \overline{b}) + \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
    2. The operation $+$ is commutative: $\overline{a} + \overline{b} = \overline{b} + \overline{a}$ for any $\overline{a}$ and $\overline{b}$.
    3. The residue class $\overline{0}$ is an additive identity: $\overline{a} + \overline{0} = \overline{a}$ for any $\overline{a}$.
    4. Every residue class $\overline{a}$ has an additive inverse $-\overline{a}$ satisfying $\overline{a} + (-\overline{a}) = \overline{0}$.
    5. The operation $\cdot$ is associative: $\overline{a} \cdot (\overline{b} \cdot \overline{c}) = (\overline{a} \cdot \overline{b}) \cdot \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
    6. The operation $\cdot$ is commutative: $\overline{a} \cdot \overline{b} = \overline{b} \cdot \overline{a}$ for any $\overline{a}$ and $\overline{b}$.
    7. The operation $\cdot$ distributes over $+$: $\overline{a} \cdot (\overline{b} + \overline{c}) = \overline{a} \cdot \overline{b} + \overline{a} \cdot \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
    8. The residue class $\overline{1}$ is a multiplicative identity: $\overline{1} \cdot \overline{a} = \overline{a}$ for any $\overline{a}$.
        - Proof: All of these follow immediately from the corresponding properties of arithmetic in $F[x]$.
- Like in $\mathbb{Z}$, we can (and will!) abuse notation and drop the bar notation, with the understanding that all of our calculations are to be considered "modulo p".
- When F is an infinite field, there will be infinitely many residue classes in $F[x]/p$, so we cannot sensibly write down addition and multiplication tables. However, we can certainly construct such tables when F is finite.
- Example: Here are the addition and multiplication tables for $\mathbb{F}_2[x]/p$ 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

    - As an example of the computations, we have $x + (x + 1) = 2x + 1 = 1$ since $2 = 0$ in $\mathbb{F}_2$, and also $x \cdot (x + 1) = x^2 + x \equiv (x + 1) + x = 1$ (because $x^2 \equiv x + 1$ which follows from rearranging the statement $p \equiv 0 \bmod p$).
    - We can see that $\mathbb{F}_2[x]/p$ is a field in this case, since every nonzero residue class has a multiplicative inverse.
    - Observe also that the polynomial $p(x) = x^2 + x + 1$ is irreducible in $\mathbb{F}_2[x]$, since it has degree 2 but no roots.
- Example: Here are the addition and multiplication tables for $\mathbb{F}_2[x]/q$ with $q = x^2$:

$+$	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

    - We can see that $\mathbb{F}_2[x]/q$ is not a field, because the element x does not have a multiplicative inverse (although 1 and $x + 1$ do).

- Notice that $q(x) = x^2$ is reducible in $\mathbb{F}_2[x]$, and has the factorization $x^2 = x \cdot x$.

- Example: Here is the multiplication table for $\mathbb{F}_3[x]/r$ with $r = x^2 + 1$:

$\cdot$	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	$2x+1$	$x+1$	$2x+2$	$2x+1$	$x+2$
$2x+1$	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	$2x$

  - Notice that $\mathbb{F}_3[x]/r$ is a field, since every nonzero residue class is invertible.
  - Observe also that the polynomial $p(x) = x^2 + 1$ is irreducible in $\mathbb{F}_3[x]$, since it has degree 2 but no roots.

- As suggested by the examples above (and by the analogy between $\mathbb{Z}$ and $F[x]$), we can characterize the invertible classes in $F[x]/p$:

- Theorem (Invertible Elements in $F[x]/p$): Let F be a field and $p \in F[x]$ be nonzero. Then the residue class $\overline{r}$ in $F[x]/p$ has a multiplicative inverse if and only if r and p are relatively prime.

  - Proof: First suppose that r and p are relatively prime. Then by the Euclidean algorithm, we can write $1 = c_r r + c_p p$ for some polynomials c_r, c_p. Then $\overline{c_r} \cdot \overline{r} = \overline{1}$, meaning that $\overline{r}$ is invertible in $F[x]/p$.
  - Conversely, suppose that $\overline{r}$ is invertible in $F[x]/p$ with multiplicative inverse $\overline{c_r}$. Then $\overline{c_r} \cdot \overline{r} = \overline{1}$ so that $c_r r \equiv 1 \pmod{p}$, meaning that there exists some polynomial c_p with $c_r r + c_p p = 1$. But any common divisor of r and p must then divide $c_r r + c_p p = 1$, and thus we see that r and p are relatively prime.

- Per the proof of the theorem above, we can use the Euclidean algorithm in $F[x]$ to compute multiplicative inverses when they exist:

- Example: Find the multiplicative inverse of $x^2 + 2$ in $\mathbb{F}_5[x]$ modulo $x^3 + 1$.

  - First we apply the Euclidean algorithm in $\mathbb{F}_5[x]$:

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

    and so the gcd of $x^2 + 2$ and $x^3 + 1$ is 1.
  - By back-solving, we obtain

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

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

- In analogy with the fact that $\mathbb{Z}/m\mathbb{Z}$ is a field precisely when m is prime, we also see that $F[x]/p$ is a field precisely when p is irreducible:

- Corollary: Let F be a field and $p \in F[x]$ have positive degree. Then $F[x]/p$ is a field if and only if p is irreducible.

  - Proof: By the previous theorem, we see that if p is irreducible then every nonzero residue class modulo p is invertible. Furthermore, if $\deg(p) > 0$, then $\overline{1} \neq \overline{0}$, so $F[x]/p$ is a field.
  - Inversely, if p is reducible, then (again as above) there are non-invertible residue classes in $F[x]/p$.

- By finding irreducible polynomials in $\mathbb{F}_p[x]$, we can use the corollary above to construct additional finite fields.
- Example: Construct a finite field with 27 elements.
  - Since $27 = 3^3$, we can construct a finite field with 27 elements as $\mathbb{F}_3[x]/p$ where p is an irreducible polynomial of degree 3.
  - One possible choice is the polynomial $p(x) = x^3 + 2x + 1$: it has no roots, since $p(0) = p(1) = p(2) = 1$, so (since it has degree 3) it is irreducible.
  - Therefore, $\mathbb{F}_3[x]/p$ is a field with $3^3 = 27$ elements, as required.

1.3 Survey of Rings

- Many of the properties of $\mathbb{Z}$, $\mathbb{Z}/m\mathbb{Z}$, $F[x]$, and $F[x]/p$ can be abstracted and generalized to a broader class of objects known as rings. In this section we give a brief survey of some basic aspects of ring theory, with the ultimate aim to illustrate how to perform and compute with "modular arithmetic" in general rings.

1.3.1 The Formal Definition of a Ring

- Definition: A ring is any set R having two (closed) binary operations $+$ and $\cdot$ that satisfy the six axioms [R1]-[R6]:

  [R1] The operation $+$ is associative: $a + (b + c) = (a + b) + c$ for any elements a, b, c in R.

  [R2] The operation $+$ is commutative: $a + b = b + a$ for any elements a, b in R.

  [R3] There is an additive identity 0 satisfying $a + 0 = a$ for all a in R.

  [R4] Every element a has an additive inverse $-a$ satisfying $a + (-a) = 0$.

  [R5] The operation $\cdot$ is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for any elements a, b, c in R.

  [R6] The operation $\cdot$ distributes over $+$: $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$ for any elements a, b, c in R.

- Certain rings will also possess additional properties, which arise often enough that we give them names:
- Definition: If a ring satisfies axiom [R7], we say it is a commutative ring.

  [R7] The operation $\cdot$ is commutative: $a \cdot b = b \cdot a$ for any elements a, b in R.

- Definition: If a ring satisfies axiom [R8], we say it is a ring with identity (or a "ring with 1").

  [R8] There is a multiplicative identity $1 \neq 0$, satisfying $1 \cdot a = a = a \cdot 1$ for all a in R.

- Definition: If a ring with identity further satisfies the axiom [D], it is called a division ring. A commutative division ring is called a field.

  [D] Every nonzero a in R has a multiplicative inverse a^{-1} satisfying $a \cdot a^{-1} = 1 = a^{-1} \cdot a$.

1.3.2 Examples of Rings

- Here is a list of examples (and non-examples) of rings[5]:
- Example: The integers $\mathbb{Z}$ are a commutative ring with identity.
- Example: The set of even integers is a commutative ring that does not have an identity.
  - The properties [R1]-[R7] all follow from their counterparts in $\mathbb{Z}$: [R3] follows because 0 is an even integer, and [R4] follows because n is an even integer if and only if $-n$ is an even integer.

[5] For brevity, when we do not specify the operations $+$ and $\cdot$, they are always assumed to be the standard addition and multiplication operations on the corresponding sets.

- This ring does not have a multiplicative identity because there is no solution to $2n = 2$ inside the set of even integers.

- Non-Example: The set of odd integers is not a ring.
  - The problem is that, although multiplication of two odd integers does return an odd integer, the sum of two odd integers is not odd: thus, the operation $+$ is not defined on the set of odd integers.

- Example: The set $\mathbb{Z}/m\mathbb{Z}$ of residue classes modulo m form a commutative ring with identity.
  - Furthermore, if p is a prime, we know that all of the nonzero residue classes modulo p are invertible, meaning that $\mathbb{Z}/p\mathbb{Z}$ is a field.
  - Indeed, the only residue classes that are invertible modulo m are those relatively prime to m, so if m is not prime, then $\mathbb{Z}/m\mathbb{Z}$ is not a field.

- Example: The rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are all examples of fields.
  - Recall that $\mathbb{C}$ is the set of numbers of the form $a + bi$, where a and b are real numbers and $i^2 = -1$.
  - Addition and multiplication in $\mathbb{C}$ are as follows: $(a+bi)+(c+di) = (a+c)+(b+d)i$, and $(a+bi)\cdot(c+di) = ac + adi + bci + bdi^2 = (ac - bd) + (ad + bc)i$.

- Example: If F is a field, the set $F[x]$ of polynomials in x with coefficients from F forms a commutative ring with identity.
  - More generally, if R is any ring, we can consider the ring $R[x]$ of polynomials with coefficients from R (we have already implicitly done this when discussing polynomials with integer coefficients).
  - Warning: When R is not commutative, the polynomial ring $R[x]$ can have unintuitive properties. Even the case where R is commutative can carry complications, if (for example) R possesses zero divisors, as there are examples where factorizations are not unique.

- Example: The set of complex numbers of the form $a + bi$ where $a, b \in \mathbb{Z}$ are a commutative ring with identity.
  - This ring is denoted $\mathbb{Z}[i]$ (read as: "$\mathbb{Z}$ adjoin i") and is also often called the Gaussian integers.
  - The properties [R1]-[R8] all follow from their counterparts in $\mathbb{C}$: [R3] follows because $0 = 0 + 0i$, and [R4] follows because we have $-(a + bi) = (-a) + (-b)i$.

- Example: The set of real numbers of the form $a + b\sqrt{2}$ where $a, b \in \mathbb{Z}$ are a commutative ring with identity.
  - This ring is denoted $\mathbb{Z}[\sqrt{2}]$. The addition and multiplication are defined in a similar way as for the complex numbers and Gaussian integers: $(a+b\sqrt{2})+(c+d\sqrt{2}) = (a+c)+(b+d)\sqrt{2}$, and $(a+b\sqrt{2})\cdot(c+d\sqrt{2}) = ac + ad\sqrt{2} + bc\sqrt{2} + 2bd = (ac + 2bd) + (ad + bc)\sqrt{2}$.

- We can also see the structure of a ring in collections of functions:

- Example: If S is any set and A is any ring, the collection R of functions $f : S \to A$, with operations $(f + g)(x) = f(x) + g(x)$ and $(f \cdot g)(x) = f(x)g(x)$, forms a ring.
  - Thus, for example, if A is the set of real numbers, with $f(x) = x^2$ and $g(x) = 3x^2$, then $(f+g)(x) = 4x^2$ and $(fg)(x) = 3x^4$.
  - Ultimately, each of the properties [R1]-[R6] follows from the corresponding property of A. The additive identity is the "identically-zero function" 0_S that is 0 on each element of S, and the additive inverse $-f$ of f is defined as $(-f)(x) = -f(x)$ for each x in S.
  - If A is commutative, then it is easy to see that R will also be commutative. Likewise, if A has a 1, then the "identically-1 function" 1_S that is 1 on each element of S, is a multiplicative identity in R.

- Example: The collection R of continuous real-valued functions $f : \mathbb{R} \to \mathbb{R}$, with operations $(f + g)(x) = f(x) + g(x)$ and $(f \cdot g)(x) = f(x)g(x)$, forms a commutative ring.

- The operations are well-defined because the sum and product of two continuous functions is continuous.
- The remaining properties [R1]-[R6] follows from the same observations as in the example above.

- So far, all of the rings we have listed are commutative. Here are a few that are not:

- Example: The set of 2×2 matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with real number entries, denoted $M_{2\times 2}(\mathbb{R})$, forms a noncommutative ring with identity.
  - Explicitly, the operations in this ring are $\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$ and $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$.
  - It is a straightforward algebraic computation to verify axioms [R1]-[R6]: the additive identity is the zero matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, and the additive inverse of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is of course $\begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}$.
  - The multiplicative identity is the famous "identity matrix" $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
  - However, the 2×2 matrices are not a commutative ring, since (for example) we have $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ while $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.
  - Remark: More generally, for any integer $n \geq 2$, the set of $n \times n$ matrices with entries from any field F, denoted $M_{n\times n}(F)$, forms a noncommutative ring with identity.

- Example: The set $\mathbb{H}$ of real quaternions $a + bi + cj + dk$, for real numbers a, b, c, d and "imaginary units" i, j, k satisfying the relations $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, and $ki = -ik = j$, form a noncommutative ring with identity.
  - This ring was first characterized by William Rowan Hamilton in 1843 (whence the name $\mathbb{H}$), and is, historically speaking, one of the first examples of a noncommutative ring.
  - The addition and multiplication operations are defined similarly to those in the complex numbers: addition works "componentwise", so that $(a + bi + cj + dk) + (a' + b'i + c'j + d'k) = (a + a') + (b + b')i + (c + c')j + (d + d')k$.
  - Multiplication is defined using the distributive law and the relations listed above, taking care to keep the terms in the proper order when multiplying. (The real number coefficients commute with the "imaginary units" i, j, and k.)
  - Thus, for example, we have

$$\begin{aligned} (1 + i - k) \cdot (2 + 3i + j) &= (1 + i - k) \cdot 2 + (1 + i - k) \cdot 3i + (1 + i - k) \cdot j \\ &= (2 + 2i - 2k) + (3i - 3 - 3j) + (j + k + i) \\ &= -1 + 6i - 2j - k. \end{aligned}$$

  - In fact, the real quaternions are a division ring: one may verify that $(a+bi+cj+dk)(a-bi-cj-dk) = a^2+b^2+c^2+d^2$, and so the nonzero quaternion $a+bi+cj+dk$ has a multiplicative inverse $\dfrac{a-bi-cj-dk}{a^2+b^2+c^2+d^2}$.
  - The quaternions originally arose in the study of 3-dimensional geometry. As a hint of this connection, we will note that under the standard notation for the coordinate vectors in 3-space, namely with $\mathbf{i} = \langle 1, 0, 0\rangle$, $\mathbf{j} = \langle 0, 1, 0\rangle$, and $\mathbf{k} = \langle 0, 0, 1\rangle$, then $\mathbf{i} \times \mathbf{j} = \mathbf{k} = -\mathbf{j} \times \mathbf{i}$, and similarly for the other possible cross products.

- Example: If V is a vector space of dimension larger than 1, the set $\mathcal{L}(V, V)$ of linear transformations from V to V is a noncommutative ring with 1 under the operations of function addition and function composition: $(S + T)(\mathbf{v}) = S\mathbf{v} + T\mathbf{v}$ and $(ST)\mathbf{v} = S(T\mathbf{v})$.

- This ring is not commutative because $ST \neq TS$ in general (since linear transformations generally do not commute with one another, in the same way that matrices do not). The multiplicative identity is the "identity transformation" with $I(\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v}$ in V.

- As a final observation, we remark that if we have a set with an addition operation, we can make it into a ring in a trivial way. Two examples are as follows:

- Example: If S is $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, with $+$ taken to be normal addition, and $\cdot$ defined so that $a \cdot b = 0$ for every a and b, then S is a commutative ring.

  - All of the multiplicative axioms immediately reduce to the true statement $0 = 0$. Of course, this ring has no multiplicative identity.

- Example: The set $R = \{0\}$, with operations $0 + 0 = 0$ and $0 \cdot 0 = 0$, is a commutative ring.

  - All of the axioms follow trivially. In fact, this ring even has a multiplicative identity! (But it is not a ring with 1 because we require $1 \neq 0$.)
  - This ring is known as the trivial ring, and is the only ring where $1 = 0$.

1.3.3 Basic Properties of Rings

- Our immediate goal in discussing rings is to study properties of arithmetic in $\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z}$ that generalize to arbitrary rings. To this end, we begin by establishing a number of basic properties of ring arithmetic.

  - As in $\mathbb{Z}$, we define the binary operation of subtraction by setting $a - b = a + (-b)$. We also often use implicit multiplication, and drop the $\cdot$ notation.
  - We can define scaling of a ring element a by a positive integer as repeated addition: $na = \underbrace{a + a + a + \cdots + a}_{n \text{ terms}}$.

    By associativity of addition, this notation is well-defined. In a ring with 1, this notation coincides with the product of ring elements $n \cdot a$, but (as we would desire) it is true that $n \cdot a = na$.
  - We can also define exponentiation of a ring element a as $a^k = \underbrace{a \cdot a \cdot a \cdot \cdots \cdot a}_{k \text{ terms}}$, for any positive integer k.

    By associativity of multiplication, this notation is well-defined.

- Proposition (Basic Arithmetic): Let R be any ring. The following properties hold in R:

  1. The additive identity 0 is unique, as is the multiplicative identity 1 (if R has a 1).
  2. Addition has a cancellation law: for any $a, b, c \in R$, if $a + b = a + c$, then $b = c$.
  3. Additive inverses are unique.
  4. For any $a \in R$, $0 \cdot a = 0 = a \cdot 0$.
  5. For any $a \in R$, $-(-a) = a$.
  6. For any $a \in R$, $(-1) \cdot a = -a = a \cdot (-1)$.
  7. For any $a, b \in R$, $-(a + b) = (-a) + (-b)$.
  8. For any $a, b \in R$, $(-a) \cdot b = -(a \cdot b) = a \cdot (-b)$, and $(-a) \cdot (-b) = a \cdot b$.
  9. For any positive integers m and n and any $a \in R$, $ma + na = (m + n)a$, $m(na) = (mn)a$, $a^{m+n} = a^m a^n$, and $a^{mn} = (a^m)^n$.
     - Each of these is a fairly straightforward calculation from the ring axioms.

- An important property of $\mathbb{Z}$ that does *not* hold in general rings is the statement that $ab = 0$ implies $a = 0$ or $b = 0$.

  - Indeed, we have already seen examples of situations in $\mathbb{Z}/m\mathbb{Z}$ where m is not prime where $ab = 0$ but $a, b \neq 0$: for example, in $\mathbb{Z}/6\mathbb{Z}$, we have the equality $\overline{2} \cdot \overline{3} = \overline{0}$.

- Inversely, it is also possible for a general ring to contain many elements that have multiplicative inverses (unlike in $\mathbb{Z}$, where the only elements with multiplicative inverses are 1 and -1).
- Definition: In a ring R, we say that an element a is a zero divisor if $a \neq 0$ and there exists a nonzero $b \in R$ such that $ab = 0$ or $ba = 0$. (Note in particular that 0 is *not* a zero divisor!)
- Definition: In a ring R with $1 \neq 0$, we say that an element a is a unit if there exists a $b \in R$ such that $ab = 1 = ba$. The set of units in R is denoted $R^\times$.
    - Example: In $\mathbb{Z}$, there are no zero divisors, and the units are ± 1.
    - Example: In $\mathbb{Z}/m\mathbb{Z}$, the units are the residue classes relatively prime to m, while the zero divisors are the nonzero classes having a nontrivial common divisor with m. In particular, every nonzero residue is either a unit or a zero divisor.
    - Example: In a field, every nonzero element is a unit. Indeed, a commutative ring with 1 is a field precisely when every nonzero element is a unit.
    - Example: In the ring $\mathbb{Z}[\sqrt{2}]$, the integers 1 and -1 are units, but the element $\sqrt{2}+1$ is also a unit, because $(\sqrt{2}+1)\cdot(\sqrt{2}-1) = 1$. Note that $\mathbb{Z}[\sqrt{2}]$ is not a field, however, because $\sqrt{2}$ is not a unit.
- Here are a few basic properties of units and zero divisors:
- Proposition (Units and Zero Divisors): Let R be a ring with $1 \neq 0$.
    1. The multiplicative inverse of a unit is unique.
        - Proof: If a is a unit with $ab = 1 = ba$ and also $ac = 1 = ca$, then $b = b(ac) = (ba)c = c$.
    2. The product of two units is a unit, as is the multiplicative inverse of a unit.
        - Proof: If a is a unit with $ab = 1 = ba$, then by definition b is also a unit.
        - If c is another unit with $cd = 1 = dc$, then $(ac)(db) = a(cd)b = a1b = ab = 1$ and likewise $(db)(ac) = 1$ as well, so the inverse of ac is db.
    3. A unit can never be a zero divisor in R.
        - Proof: Suppose first that a is a unit and that $xa = 0$ for some $x \neq 0$.
        - Then by assumption, there is a b such that $ab = 1$, so then $x = x(ab) = (xa)b = 0b = 0$, contradicting the assumption that $x \neq 0$.
        - In the same way, if $ax = 0$ for some $x \neq 0$, then if $ba = 1$ then $x = (ba)x = b(ax) = b0 = 0$, again a contradiction.
- We give a special name to the class of commutative rings having no zero divisors, attesting to their similarity to $\mathbb{Z}$:
- Definition: A commutative ring with $1 \neq 0$ having no zero divisors is called an integral domain (or often, just a "domain"). Equivalently, R is an integral domain if R is commutative with $1 \neq 0$, and where $ab = 0$ implies $a = 0$ or $b = 0$.
    - The integers are an integral domain, as is any field.
    - More generally, any ring that is a subset of a field (such as the Gaussian integers $\mathbb{Z}[i]$) is an integral domain. In fact, the converse turns out to be true as well: any integral domain arises naturally as a subset of a field.
- Integral domains possess various fundamental properties:
- Proposition (Cancellation in Domains): Suppose R is an integral domain. Then multiplication in R has a cancellation law: if $a \neq 0$ and $ab = ac$, $b = c$.
    - Proof: Suppose that $ab = ac$: then $a(b-c) = 0$, so since R is a domain we either have $a = 0$ or $b - c = 0$. Thus, if $a \neq 0$, we have $b - c = 0$ so that $b = c$.
- Corollary: If R is a finite integral domain, then R is a field.

- Proof: Let a be any nonzero element of R, and consider the set $\{a, a^2, a^3, \ldots, a^n, \ldots\}$. Since R is finite, two of the elements of this set must be equal: say $a^j = a^{j+k}$ for some positive integers j and k.
- Then $a^j = a^{j+k}$ implies $a^j(a^k - 1) = 0$, and then since $a \neq 0$, we see $a^j \neq 0$. Thus, $a^k - 1 = 0$, so that $a \cdot a^{k-1} = 1$, meaning that a^{k-1} is the multiplicative inverse of a.

- A number of the examples of rings we described earlier arise naturally as subsets of other rings. We can easily describe this phenomenon in general:

- Definition: If R is a ring, we say a subset S of R is a subring if it also possesses the structure of a ring, under the same operations as R.

- Observe that if S is a subset of a ring, in order for the operations $+$ and $\cdot$ to be well-defined binary operations on S, it must be the case that $a + b$ and $a \cdot b$ are elements of S, for any a and b in S.
- Next, observe that axioms [R1], [R2], [R5], and [R6] in S automatically follow from the corresponding properties of R.
- In order for [R3] to hold, we must have an additive identity 0_S in S with the property that $a + 0_S = a$ for every a in S. However, by the additive cancellation law in R, since $a + 0_R = a = a + 0_S$, we see that $0_S = 0_R$: in other words, S must contain the zero element of R.
- Finally, in order for [R4] to hold, we require that for every $a \in S$, its additive inverse $(-a)$ must also be in S.

- By employing subtraction, we can in fact combine two of these verifications:

- Proposition (Subring Criterion): A subset S of R is a subring if only if S contains the zero element of R and, for any $a, b \in S$, the elements $a - b$ and ab are also in S.

- Proof: If S is a subring, then as noted above S must contain the zero element of R and for any $a, b \in S$, we must have $a - b$ and ab in S.
- Conversely, suppose S contains 0 and that $a - b$ and ab are also in S for any $a, b \in S$. By setting $a = 0$ we see that $0 - b = -b$ is in S, and then by setting $b = -c$ we see that $a - (-c) = a + c$ is in S.
- Therefore, S contains 0 and is closed under addition, multiplication, and taking additive inverses. By the observations above, S is therefore a subring.

- Using the subring criterion, we can construct many more examples of rings.

- Example: $\mathbb{Z}$ is a subring of $\mathbb{Q}$, which is a subring of $\mathbb{R}$, which is a subring of $\mathbb{C}$.

- Example: The trivial ring $\{0\}$ is a subring of any ring.

- Example: The even integers $2\mathbb{Z}$ are a subring of $\mathbb{Z}$, as are (more generally) the integer multiples of n, written $n\mathbb{Z}$.

- In fact, every subring of $\mathbb{Z}$ is of the form $n\mathbb{Z}$ for some integer n (the "trivial" subring $\{0\}$ corresponds to the case $n = 0$).
- To see this, suppose S is a subring of $\mathbb{Z}$, and let T be the set of positive elements in this subring. If T is empty, then $S = \{0\}$, and otherwise, T must have a minimal element n by the well-ordering principle. Then S contains $n\mathbb{Z}$.
- We claim any element of S must be a multiple of n, so that $S = n\mathbb{Z}$: by the division algorithm, if S contained an integer not divisible by n, the remainder upon dividing a by n would be a positive element of S smaller than n, contradiction. Thus, $S = n\mathbb{Z}$.

- Example: The set of rational numbers having denominator equal to a power of 2 (i.e., that are of the form $n/2^k$ for an integer n and nonnegative integer k), forms a subring of $\mathbb{Q}$.

- Example: The set of upper-triangular 2×2 matrices with real entries (i.e., those of the form $\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$) forms a subring of $M_{2\times 2}(\mathbb{R})$.

- Example: The set of differentiable real-valued functions is a subring of the ring of continuous real-valued functions, which is in turn a subring of the ring of all real-valued functions.
- We can also construct new rings using Cartesian products.
  - Recall that if S and T are sets, the Cartesian product $S \times T$ is the set of ordered pairs (s, t) where $s \in S$ and $t \in T$.
- Proposition (Cartesian Products of Rings): If A and B are rings, then the Cartesian product $A \times B$ is also a ring, with operations performed componentwise: $(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$ and $(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2, b_1 b_2)$.
  - Proof: Each of the properties [R1]-[R6] follows from the corresponding properties of A and B. The additive identity is $(0, 0)$, and additive inverses are given by $-(a, b) = (-a, -b)$.
  - Note that if A and B are commutative, then so is $A \times B$; likewise, if A and B have a 1, then $(1_A, 1_B)$ is the multiplicative identity in $A \times B$.
- Example: The ring $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ is a commutative ring with 1.
  - This ring has six elements: $(0, 0)$, $(0, 1)$, $(0, 2)$, $(1, 0)$, $(1, 1)$, and $(1, 2)$. The additive identity is $(0, 0)$ and the multiplicative identity is $(1, 1)$.
  - Operations are performed "modulo 2" in the first coordinate and "modulo 3" in the second coordinate, so for example we have $(1, 1) + (0, 2) = (1, 0)$ and $(1, 2) \cdot (0, 2) = (0, 1)$.

1.3.4 Ideals

- Our next task is to generalize the idea of "modular arithmetic" into general rings.
  - To motivate our discussion, recall the ideas behind the construction of $\mathbb{Z}/m\mathbb{Z}$ and R/pR where $R = F[x]$: we first defined modular modular congruences and studied their properties, and then we constructed residue classes and showed that the collection of all residue classes had a ring structure.
- In both $\mathbb{Z}$ and $F[x]$, we defined modular congruences using divisibility, but let us take a broader approach: if I is a subset of R (whose properties we intend to characterize in a moment) let us say that two elements $a, b \in R$ are "congruent modulo I" if $a - b \in I$.
  - This is a generalization of both types of congruence we have described thus far: for $\mathbb{Z}/m\mathbb{Z}$, the set I consists of the multiples of m, while for R/pR, the set I consists of the multiples of p.
  - We would like "congruence modulo I" to be an equivalence relation: this requires $a \equiv a \pmod{I}$, $a \equiv b \pmod{I}$ implies $b \equiv a \pmod{I}$, and $a \equiv b \pmod{I}$ and $b \equiv c \pmod{I}$ implies $a \equiv c \pmod{I}$.
  - It is easy to see that these three conditions require $0 \in I$, that I be closed under additive inverses, and that I be closed under addition. (Thus, I is in fact closed under subtraction.)
  - Furthermore, we would like the congruences to respect addition and multiplication: if $a \equiv b \pmod{I}$ and $c \equiv d \pmod{I}$, then we want $a + c \equiv b + d \pmod{I}$ and $ac \equiv bd \pmod{I}$.
  - In terms of ring elements, this is equivalent to the following: if $b = a + r$ and $d = c + s$ for some $r, s \in I$, then we want $(b+d)-(a+c) = r+s$ to be in I, and we also want $bd-ac = (a+r)(c+s)-ac = as+rc+rs$ to be in I.
  - The first condition clearly follows from the requirement that I is closed under addition. It is a bit less obvious how to handle the second condition, but one immediate implication follows by setting $a = c = 0$: namely, that $rs \in I$.
  - Thus, I must be closed under multiplication, so it is in fact a subring of R.
  - But the well-definedness of multiplication actually requires more: since $0 \in I$, we can set $r = 0$ to see that $as \in I$, and we can also set $s = 0$ to see that $rc \in I$.
  - So in fact, I must be closed under (left and right) multiplication by *arbitrary* elements of R, in addition to being a subring. It is then easy to see that this condition is also sufficient to ensure that $a \equiv b \pmod{I}$ and $c \equiv d \pmod{I}$ imply $a + c \equiv b + d \pmod{I}$ and $ac \equiv bd \pmod{I}$.

- Our last task is to define residue classes and then the ring operations: we define the residue class $\overline{a}$ (modulo I) to be the set of ring elements b congruent to a modulo I, which is to say, $\overline{a} = \{a + r : r \in I\}$.
- Then we take the operations on residue classes to be $\overline{a} + \overline{b} = \overline{a+b}$ and $\overline{a} \cdot \overline{b} = \overline{a \cdot b}$: then from our properties of congruences, we can verify that these operations are well-defined and that the collection of residue classes forms a ring.

- Now that we have established the basic properties of the classes of the sets I we can use to construct congruences, we can run through the discussion more formally.
- Definition: A subset I of a ring R that is closed under arbitrary left and right multiplication by elements of R is called an ideal of R (or, for emphasis, a two-sided ideal).
  - Explicitly, I is an ideal if I contains 0 and for any $x, y \in I$ and any $r \in R$, the elements $x - y$, rx, and xr are all in I.
  - There are "one-sided" notions of ideals as well (a left ideal is closed under arbitrary left multiplication, while a right ideal is closed under arbitrary right multiplication). If R is commutative, then left ideals, right ideals, and two-sided ideals are the same.
- Here are a few basic examples of ideals:
  - Example: The subrings $n\mathbb{Z}$ are ideals of $\mathbb{Z}$, since they are clearly closed under arbitrary multiplication by elements of $\mathbb{Z}$.
  - Example: If $R = F[x]$ and p is any polynomial, the subring pR of multiples of p is an ideal of $F[x]$, since it is closed under arbitrary multiplication by polynomials in $F[x]$.
  - Non-example: The subring $\mathbb{Z}$ of $\mathbb{Q}$ is not an ideal of $\mathbb{Q}$, since it is not closed under arbitrary multiplication by elements of $\mathbb{Q}$, since for example if we take $r = \frac{1}{3} \in \mathbb{Q}$ and $x = 4 \in \mathbb{Z}$, the element $rx = \frac{4}{3}$ is not in $\mathbb{Z}$.
  - Example: For any ring R, the subrings $\{0\}$ and R are ideals of R. We refer to $\{0\}$ as the trivial ideal (or the "zero ideal") and refer to any ideal $I \neq R$ as a proper ideal (since it is a proper subset of R).
- Here are a few more examples (and non-examples) of ideals.
- Example: In the polynomial ring $\mathbb{Z}[x]$, determine whether the set S of polynomials with even constant term (i.e., the polynomials of the form $2a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ for integers a_i) forms an ideal.
  - It is easy to see that $0 \in S$ and that S is closed under subtraction.
  - Furthermore, if $q(x)$ is any other polynomial, and $p(x) \in S$, then $p(x)q(x)$ also has even constant term, so it is also in S.
  - Thus, S is closed under multiplication by elements of $\mathbb{Z}[x]$, so it $\boxed{\text{is an ideal}}$.
- Example: Determine whether the set $S = \{\overline{0}, \overline{2}, \overline{4}, \overline{6}\}$ of "even" residue classes is an ideal of $\mathbb{Z}/8\mathbb{Z}$.
  - We have $0 \in S$, and it is a straightforward calculation to see that S is closed under subtraction, since the sum of two "even" residue classes modulo 8 will still be even.
  - Furthermore, the product of any residue class with an even residue class will again be an even residue class (since 8 is even), so S is closed under multiplication by arbitrary elements of R. Thus, S $\boxed{\text{is an ideal}}$.
- Example: Determine whether the set $S = \{(2a, 3a) : a \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z} \times \mathbb{Z}$.
  - We have $0 \in S$, and $(2a, 3a) - (2b, 3b) = (2(a-b), 3(a-b))$ so S is closed under subtraction.
  - But, for example, we can see that $(1, 2) \cdot (2, 3) = (2, 6)$ is not in S, even though $(2, 3)$ is, so S is not closed under arbitrary multiplication by elements of $\mathbb{Z} \times \mathbb{Z}$. Thus, S $\boxed{\text{is not an ideal}}$.
- Several of the examples above are particular instances of a general class of ideals:
- Proposition (Principal Ideals): If R is a commutative ring with 1, the set $(a) = \{ra : r \in R\}$ of all R-multiples of a forms a (two-sided) ideal of R, known as the principal ideal generated by a.

- Proof: Since $0a = 0$ we see $0 \in (a)$. Furthermore, since $ra - sa = (r-s)a$ we see that (a) is closed under subtraction.
- Furthermore, if $t \in R$ then we have $t(ra) = (tr)a$, so since R is commutative, (a) is closed under multiplication by arbitrary elements of R. Thus, (a) is an ideal.

1.3.5 Quotient Rings

- Now that we have discussed ideals, we can use them to study residue classes, and thereby discuss construct "quotient rings".
- Definition: If I is an ideal of the ring R, then we say a is congruent to b modulo I, written $a \equiv b \pmod{I}$, if $a - b \in I$.
  - As in $\mathbb{Z}$ and $F[x]$, congruence modulo I is an equivalence relation that respects addition and multiplication. The proofs are the same as in $\mathbb{Z}$ and $F[x]$, once we make the appropriate translations from "divisibility" to "containment in I".
- Proposition (Ideal Congruences): Let I be an ideal of R and $a, b, c, d \in R$. Then the following are true:
  1. $a \equiv a \pmod{I}$.
  2. $a \equiv b \pmod{I}$ if and only if $b \equiv a \pmod{I}$.
  3. If $a \equiv b \pmod{I}$ and $b \equiv c \pmod{I}$, then $a \equiv c \pmod{I}$.
  4. If $a \equiv b \pmod{I}$ and $c \equiv d \pmod{I}$, then $a + c \equiv b + d \pmod{I}$.
  5. If $a \equiv b \pmod{I}$ and $c \equiv d \pmod{I}$, then $ac \equiv bd \pmod{I}$.
     - Each of these is a straightforward calculation using the definition of an ideal.
- Now we can define residue classes:
- Definition: If I is an ideal of the ring R, then for any $a \in R$ we define the residue class of a modulo I to be the set $\overline{a} = a + I = \{a + x : x \in I\}$. This set is also called the coset of I represented by a.
  - We will use the notation $\overline{a}$ and $a+I$ interchangeably. (The latter is intended to evoke the idea of "adding" a to the set I.)
  - We observe, as with our previous examples of residue classes, that any two residue classes are either disjoint or identical and that they partition R: specifically, $\overline{a} = \overline{b}$ if and only if $a \equiv b \pmod{I}$ if and only if $a - b \in I$.
- All that remains is to verify that the residue classes form a ring, in the same way as in $\mathbb{Z}$ and $F[x]$:
- Theorem (Quotient Rings): Let I be an ideal of the ring R. Then the collection of residue classes modulo I forms a ring, denoted R/I (read as "R mod I"), under the operations $\overline{a} + \overline{b} = \overline{a+b}$ and $\overline{a} \cdot \overline{b} = \overline{ab}$. (This ring is called the quotient ring of R by I.) If R is commutative then so is R/I, and likewise if R has a 1 then so does R/I.
  - Remark: The notation R/I is intended to emphasize the idea that I represents a single element (namely, $\overline{0}$) in the quotient ring R/I, and the other elements in R/I are "translates" of I. In this way, R/I is the ring obtained from R by "collapsing" or "dividing out" by I, whence the name "quotient ring".
  - The proof of this fact is exactly the same as in the cases of $\mathbb{Z}$ and $F[x]$, and only requires showing that the operations are well-defined.
  - Proof: First we must show that the addition and multiplication operations are well-defined: that is, if we choose different elements $a' \in \bar{a}$ and $b' \in \bar{b}$, the residue class of $a' + b'$ is the same as that of $a + b$, and similarly for the product.
  - To see this, if $a' \in \bar{a}$ then $a' \equiv a \pmod{I}$, and similarly if $b' \in \bar{b}$ then $b' \equiv b \pmod{I}$.
  - Then $a' + b' \equiv a + b \pmod{I}$, so $\overline{a' + b'} = \overline{a + b}$. Likewise, $a'b' \equiv ab \pmod{I}$, so $\overline{a'b'} = \overline{ab}$.
  - Thus, the operations are well-defined.

- For the ring axioms [R1]-[R6], we observe that associativity, commutativity, and the distributive laws follow immediately from the corresponding properties in R: the additive identity in R/I is $\overline{0}$ and the additive inverse of $\overline{a}$ is $\overline{-a}$.
- Finally, if R is commutative then so will be the multiplication of the residue classes, and if R has a 1 then the residue class $\overline{1}$ is easily seen to be a multiplicative identity in R/I.

- This general description of "quotient rings" generalizes the two examples we have previously discussed: $\mathbb{Z}/m\mathbb{Z}$ and $F[x]/p$.
  - To be explicit, $\mathbb{Z}/m\mathbb{Z}$ is the quotient of $\mathbb{Z}$ by the ideal $m\mathbb{Z}$, while $F[x]/p$ is the quotient of the polynomial ring $F[x]$ by the principal ideal (p) consisting of all multiples of p.
  - It is not hard to see that the integer congruence $a \equiv b \pmod{m}$, which we originally defined as being equivalent to the statement $m|(b-a)$, is the same as the congruence $a \equiv b \pmod{I}$ where I is the ideal $m\mathbb{Z}$, since $b - a \in m\mathbb{Z}$ precisely when $b - a$ is a multiple of m.

- Here are some additional examples of quotient rings:

- Example: If R is any ring, the quotient ring of R by the zero ideal, namely $R/0$, has the same structure as R itself, while the quotient ring of R by itself, namely R/R, has the same structure as the trivial ring $\{0\}$.

- Example: In $R = \mathbb{Z}[x]$, with I consisting of all multiples of $x^2 + 1$, describe the structure of the quotient ring R/I.
  - It is easy to see that I is an ideal of R, since it is a subring that is closed under arbitrary multiplication by elements of R.
  - From our discussion of polynomial rings, we know that the residue classes in R/I are represented uniquely by residue classes of the form $\overline{a + bx}$ where $a, b \in \mathbb{Z}$. Note that in this quotient ring, we have $\overline{x}^2 + \overline{1} = \overline{0}$, which is to say, $\overline{x}^2 = -\overline{1}$.
  - The addition in this quotient ring is given by $\overline{a + bx} + \overline{c + dx} = \overline{(a + c) + (b + d)x}$ while the multiplication is given by $\overline{a + bx} \cdot \overline{c + dx} = \overline{(ac - bd) + (ad + bc)x}$, which follows from the distributive law and the fact that $\overline{x}^2 = -\overline{1}$.

- Example: In $R = \mathbb{Z}/8\mathbb{Z}$, with $I = \{0, 4\}$, describe the structure of the quotient ring R/I.
  - It is easy to see that I is an ideal of R, since it is a subring that is closed under arbitrary multiplication by elements of R. (Indeed, it is the principal ideal generated by 4.)
  - Since each residue class contains 2 elements, and R has 8 elements in total, there are four residue classes. With this observation in hand, it is not hard to give a list: $\overline{0} = I = \{0, 4\}$, $\overline{1} = 1 + I = \{1, 5\}$, $\overline{2} = 2 + I = \{2, 6\}$, and $\overline{3} = 3 + I = \{3, 7\}$.
  - Notice, for example, that in the quotient ring R/I, we have $\overline{1} + \overline{3} = \overline{0}$, $\overline{2} \cdot \overline{2} = \overline{0}$, and $\overline{2} \cdot \overline{3} = \overline{2}$: indeed, we can see that the structure of R/I is exactly the same as $\mathbb{Z}/4\mathbb{Z}$ (the labelings of the elements are even the same).

- Example: In the polynomial ring $R = \mathbb{Z}[x]$, with I consisting of the polynomials with even constant term (i.e., the polynomials of the form $2a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ for integers a_i), describe the structure of the quotient ring R/I.
  - We observe that there are only two residue classes, namely $\overline{0}$ and $\overline{1}$: to see this observe that $p(x) \in \overline{0}$ when the constant term of p is even, and $p(x) \in \overline{1}$ when the constant term of p is odd.
  - Then one can verify that the structure of this quotient ring is "the same" as $\mathbb{Z}/2\mathbb{Z}$ (with, for example, $\overline{1} + \overline{1} = \overline{0}$).

1.3.6 Ring Isomorphisms

- We have encountered several examples of rings with very similar structures.
- For example, consider the two rings $R = \mathbb{Z}/6\mathbb{Z}$ and $S = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$.
  - Here are the addition and multiplication tables in R:

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

  - Now compare those tables to the tables in S:

$+$	$(0,0)$	$(1,1)$	$(0,2)$	$(1,0)$	$(0,1)$	$(1,2)$
$(0,0)$	$(0,0)$	$(1,1)$	$(0,2)$	$(1,0)$	$(0,1)$	$(1,2)$
$(1,1)$	$(1,1)$	$(0,2)$	$(1,0)$	$(0,1)$	$(1,2)$	$(0,0)$
$(0,2)$	$(0,2)$	$(1,0)$	$(0,1)$	$(1,2)$	$(0,0)$	$(1,1)$
$(1,0)$	$(1,0)$	$(0,1)$	$(1,2)$	$(0,0)$	$(1,1)$	$(0,2)$
$(0,1)$	$(0,1)$	$(1,2)$	$(0,0)$	$(1,1)$	$(0,2)$	$(1,0)$
$(1,2)$	$(1,2)$	$(0,0)$	$(1,1)$	$(0,2)$	$(1,0)$	$(0,1)$

$\cdot$	$(0,0)$	$(1,1)$	$(0,2)$	$(1,0)$	$(0,1)$	$(1,2)$
$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
$(1,1)$	$(0,0)$	$(1,1)$	$(0,2)$	$(1,0)$	$(0,1)$	$(1,2)$
$(0,2)$	$(0,0)$	$(0,2)$	$(0,1)$	$(0,0)$	$(0,2)$	$(0,1)$
$(1,0)$	$(0,0)$	$(1,0)$	$(0,0)$	$(1,0)$	$(0,0)$	$(1,0)$
$(0,1)$	$(0,0)$	$(0,1)$	$(0,2)$	$(0,0)$	$(0,1)$	$(0,2)$
$(1,2)$	$(0,0)$	$(1,2)$	$(0,1)$	$(1,0)$	$(0,2)$	$(1,1)$

    - Notice that these tables look quite similar (especially given the artful reordering of the labels of the elements in S).
    - Indeed, if we relabel each entry n in the first set of tables with the ordered pair corresponding to its reduction modulo 2 and 3 (so that 1 becomes $(1,1)$, 2 becomes $(0,2)$, and so forth) we will obtain the second set of tables!
- For another example, consider the rings $R = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ and $S = \mathbb{F}_2[x]/(x^2 + x)$.
  - Here are the addition and multiplication tables in R:

$+$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$
$(0,0)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$
$(1,1)$	$(1,1)$	$(0,0)$	$(0,1)$	$(1,0)$
$(1,0)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$
$(0,1)$	$(0,1)$	$(1,0)$	$(1,1)$	$(0,0)$

$\cdot$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$
$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
$(1,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$
$(1,0)$	$(0,0)$	$(1,0)$	$(1,0)$	$(0,0)$
$(0,1)$	$(0,0)$	$(0,1)$	$(0,0)$	$(0,1)$

  - Now compare those tables to the tables in S:

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

    - Here, if we relabel $(0,0)$ as 0, $(1,1)$ as 1, $(1,0)$ as x, and $(0,1)$ as $x+1$, the first pair of tables becomes the second set of tables.
- Let us formalize the central idea in the examples above: in each case, we see that there is a way to "relabel" the elements of R using the elements of S in a way that preserves the ring structure.
  - The desired "relabeling" is a function $\varphi : R \to S$ with the property that φ is a bijection (so that each element of R is "labeled" with a unique element of S) and that φ respects the ring operations.
  - Explicitly, we require $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$ and $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$ for all $r_1, r_2 \in R$.
- Definition: Let R and S be rings. A ring isomorphism φ from R to S is a bijective[6] function $\varphi : R \to S$ such that $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$ and $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$ for all elements r_1 and r_2 in R.

[6] Recall that a function $\varphi : R \to S$ is injective (one-to-one) if $\varphi(x) = \varphi(y)$ implies $x = y$, and φ is surjective (onto) if for every $s \in S$ there exists an $r \in R$ with $\varphi(r) = s$. A bijective function is one that is both injective and surjective. Equivalently, φ is a bijection if it possesses a two-sided inverse function $\varphi^{-1} : S \to R$ with $\varphi(\varphi^{-1}(s)) = s$ and $\varphi^{-1}(\varphi(r)) = r$ for every $r \in R$ and $s \in S$.

- We remark here that in both of the conditions $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$ and $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$, the operations on the left are performed in R while the operations on the right are performed in S.
- Note: Isomorphisms arise in a variety of contexts (e.g., isomorphisms of vector spaces, isomorphisms of groups, etc.), and in some cases the rings we are considering may carry additional structure. We will simply say "isomorphism" when the particular type of isomorphism is clear from the context.

- Example: For $R = \mathbb{Z}/6\mathbb{Z}$ and $S = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$, the map $\varphi : R \to S$ defined via $\varphi(n \bmod 6) = (n \bmod 2, n \bmod 3)$ is an isomorphism.
  - Note that "reducing" a residue class in $\mathbb{Z}/6\mathbb{Z}$ modulo 2 or modulo 3 makes sense because 2 and 3 both divide 6, so φ is well-defined.
  - We can then appeal to the calculations above (or simply redo the calculations) to see that φ is a bijection and that $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$ and $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$ for any residue classes $r_1, r_2 \in \mathbb{Z}/6\mathbb{Z}$.

- Example: For $S = \left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \in M_{2\times 2}(\mathbb{R}) \,:\, a, b \in \mathbb{R} \right\}$, the map $\varphi : \mathbb{C} \to S$ defined via $\varphi(a+bi) = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ is an isomorphism.
  - First, we see that φ is a bijection since it has a two-sided inverse; namely, the map $\varphi^{-1} : S \to \mathbb{C}$ defined by $\varphi^{-1}\left(\begin{bmatrix} a & b \\ -b & a \end{bmatrix} \right) = a + bi$.
  - Furthermore, if $z = a + bi$ and $w = c + di$, then φ respects addition and multiplication:

$$\begin{aligned} \varphi(z+w) &= \varphi((a+c)+(b+d)i) = \begin{bmatrix} a+c & b+d \\ -(b+d) & a+c \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} + \begin{bmatrix} c & d \\ -d & c \end{bmatrix} = \varphi(z) + \varphi(w) \\ \varphi(zw) &= \varphi((ac-bd)+(ad+bc)i) = \begin{bmatrix} ac-bd & ad+bc \\ -(ad+bc) & ac-bd \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} c & d \\ -d & c \end{bmatrix} = \varphi(z) \cdot \varphi(w). \end{aligned}$$

- Definition: If there is an isomorphism $\varphi : R \to S$, we say R and S are isomorphic, and write $R \cong S$. Isomorphic rings share the same structure, except that the elements and operations may be labeled differently.

- Proposition (Properties of Isomorphisms): If R, S, T are any rings, the following hold:

1. The identity map $I : R \to R$ defined by $I(r) = r$ for all $r \in R$ is an isomorphism from R to R.
   - Proof: I is clearly a bijection and respects the ring operations.
2. If $\varphi : R \to S$ is an isomorphism, then the inverse map $\varphi^{-1} : S \to R$ is also an isomorphism.
   - Proof: Essentially by definition, φ^{-1} is also a bijection.
   - Now suppose $\varphi^{-1}(s_1) = r_1$ and $\varphi^{-1}(s_2) = r_2$, so that $\varphi(r_1) = s_1$ and $\varphi(r_2) = s_2$.
   - Then $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2) = s_1 + s_2$, meaning that $\varphi^{-1}(s_1 + s_2) = r_1 + r_2 = \varphi^{-1}(s_1) + \varphi^{-1}(s_2)$, and likewise for multiplication. Thus, φ^{-1} is also an isomorphism.
3. If $\varphi : R \to S$ and $\psi : S \to T$ are isomorphisms, then the composition $\psi\varphi : R \to T$ is also an isomorphism.
   - Proof: It is straightforward to see that the composition of two bijections is a bijection.
   - Furthermore, we have $(\psi\varphi)(r_1 + r_2) = \psi(\varphi(r_1 + r_2)) = \psi(\varphi(r_1) + \varphi(r_2)) = \psi\varphi(r_1) + \psi\varphi(r_2)$, and likewise for multiplication. Thus $\psi\varphi$ is an isomorphism.
4. If $\varphi : R \to S$ is an isomorphism, then $\varphi(0_R) = 0_S$, and if R has a 1, then so does S, and $\varphi(1_R) = 1_S$.
   - Proof: Let $s \in S$ and define $r = \varphi^{-1}(s)$. Then $s + \varphi(0_R) = \varphi(r) + \varphi(0_R) = \varphi(r + 0_R) = \varphi(r) = s$, so $\varphi(0_R)$ is an additive identity in S.
   - Similarly, if R has a 1, then $s \cdot \varphi(1_R) = \varphi(r)\varphi(1_R) = \varphi(r \cdot 1_R) = \varphi(r) = s$, so $\varphi(1_R)$ is a multiplicative identity in S.
5. If $\varphi : R \to S$ is an isomorphism, then $r \in R$ is a unit in R if and only if $\varphi(r) \in S$ is a unit in S, and if so, $\varphi(r)^{-1} = \varphi(r^{-1})$.
   - Proof: If $r \in R$ is a unit in R with inverse t, we have $1_R = rt$, so $1_S = \varphi(1_R) = \varphi(rt) = \varphi(r)\varphi(t)$ so $\varphi(r)$ is a unit in S with inverse $\varphi(t)$. The converse implication is equivalent, by (2).

6. If $\varphi : R \to S$ is an isomorphism, then R is a field if and only if S is a field.
    - Proof: Every nonzero $r \in R$ is a unit if and only if every nonzero $s \in S$ is a unit by (5), and clearly R is commutative iff S is commutative.

- Our primary interest in ring isomorphisms is that we can use them to establish that a given ring R is a field, if we can show that R is ring-isomorphic to some other field F.

1.3.7 Ring Homomorphisms

- We finally give a brief discussion of maps that respect the ring operations without the requirement that they be bijections.
- Definition: A function $\varphi : R \to S$ is a ring homomorphism if $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$ and $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$ for all elements r_1 and r_2 in R.
    - Note of course that any isomorphism is a homomorphism, but the reverse is not typically true.
- Example: If $m > 1$, show that the map $\varphi : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ defined by $\varphi(a) = \overline{a}$, so that φ maps the integer a to its associated residue class $\overline{a}$ modulo m, is a ring homomorphism.
    - From our results on residue classes, we see $\varphi(a + b) = \overline{a + b} = \overline{a} + \overline{b} = \varphi(a) + \varphi(b)$, and likewise $\varphi(a \cdot b) = \overline{a \cdot b} = \overline{a} \cdot \overline{b} = \varphi(a) \cdot \varphi(b)$. Thus, φ is a homomorphism.
    - Notice that this map is surjective but not injective (since for example $\varphi(0) = \varphi(m)$), so it is not an isomorphism.
- In essentially the same way, we see that the reduction modulo p map inside $F[x]$ is also a homomorphism:
- Example: Let F be a field with $R = F[x]$ and let $p(x) \in R$ be nonzero. Then the map $\varphi : R \to R/pR$ given by $\varphi(a) = \overline{a}$, mapping the polynomial a to its associated residue class $\overline{a}$ modulo p, is a ring homomorphism.
    - From our results on residue classes, we see $\varphi(a + b) = \overline{a + b} = \overline{a} + \overline{b} = \varphi(a) + \varphi(b)$, and likewise $\varphi(a \cdot b) = \overline{a \cdot b} = \overline{a} \cdot \overline{b} = \varphi(a) \cdot \varphi(b)$. Thus, φ is a homomorphism.
- Example: Let R be a commutative ring and $a \in R$. Show that the "evaluation at a map" $\varphi_a : R[x] \to R$ defined by $\varphi_a(p) = p(a)$ is a ring homomorphism.
    - We have $\varphi_a(p + q) = (p + q)(a) = p(a) + q(a) = \varphi_a(p) + \varphi_a(q)$ by the definition of polynomial addition.
    - Likewise, we have $\varphi_a(r_b x^b \cdot r_c x^c) = r_b r_c a^{b+c} = (r_b a^b)(r_c a^c) = \varphi_a(r_b x^b)\varphi_a(r_c x^c)$ because R is commutative.
    - Then for any polynomials p and q we see $\varphi_a(pq) = \varphi_a(p)\varphi_a(q)$ by applying distributivity and the fact that φ_a respects multiplication of individual terms and addition.
- Example: Let R and S be any rings. The "zero map" $Z : R \to S$ given by $Z(r) = 0_S$ for every $r \in R$ is a ring homomorphism.
- Example: If S is a subring of R, the map $\iota : S \to R$ given by $\iota(s) = s$ is a ring homomorphism. This map is called the inclusion map (since it simply reflects the set inclusion of S inside R).
- There exist numerous examples of maps that satisfy only one of the two requirements for being a homomorphism.
    - Non-Example: The function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 2n$ is not a homomorphism. Explicitly, although it satisfies $f(m + n) = 2(m + n) = f(m) + f(n)$, it is not multiplicative since $f(1 \cdot 1) = 2$ while $f(1) \cdot f(1) = 4$.
    - Non-Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not a homomorphism. Explicitly, although it satisfies $f(xy) = (xy)^2 = f(x)f(y)$, it is not additive since $f(1 + 1) = 4$ while $f(1) + f(1) = 2$.
- Here are a few more examples (and non-examples) of homomorphisms:

- Example: Determine whether the map $\varphi : (\mathbb{Z}/15\mathbb{Z}) \to (\mathbb{Z}/15\mathbb{Z})$ given by $\varphi(a) = 10a$ is a ring homomorphism.
  - We have $\varphi(a+b) = 10(a+b) = 10a + 10b = \varphi(a) + \varphi(b)$.
  - Likewise, $\varphi(ab) = 10ab = 100ab = (10a)(10b) = \varphi(a)\varphi(b)$, since $10 \equiv 100 \pmod{15}$.
  - Therefore, φ is a homomorphism.
- Example: Let R be the ring of infinitely differentiable real-valued functions on $\mathbb{R}$. Determine whether the derivative map $D : R \to R$ given by $D(f) = f'$ is a ring homomorphism.
  - We have $D(f+g) = (f+g)' = f' + g' = D(f) + D(g)$, so D is additive.
  - However, D does not respect ring multiplication, since for example $D(x \cdot x^2) = 3x^2$ while $D(x) \cdot D(x^2) = 2x$. Therefore, φ is not a homomorphism.
- Example: Let R be any ring. Determine whether the map $\varphi : R \to R \times R$ given by $\varphi(r) = (r, r)$ is a ring homomorphism.
  - We have $\varphi(r+s) = (r+s, r+s) = (r,r) + (s,s) = \varphi(r) + \varphi(s)$.
  - Likewise, $\varphi(rs) = (rs, rs) = (r,r)(s,s) = \varphi(r)\varphi(s)$, so φ is a homomorphism.
- Like with isomorphisms, homomorphisms have a number of basic properties.
- Proposition (Properties of Homomorphisms): If R, S, T are any rings, the following hold:
  1. If $\varphi : R \to S$ and $\psi : S \to T$ are homomorphisms, then the composition $\psi\varphi : R \to T$ is also a homomorphism.
     - Proof: Follows from the analogous calculation for isomorphisms.
  2. If $\varphi : R \to S$ is a homomorphism, then $\varphi(0_R) = 0_S$, $\varphi(-r) = -\varphi(r)$ for every $r \in R$, and $\varphi(r_1 - r_2) = \varphi(r_1) - \varphi(r_2)$ for every $r_1, r_2 \in R$.
     - Proof: For any $r \in R$, we have $\varphi(r) = \varphi(r + 0_R) = \varphi(r) + \varphi(0_R)$: thus, by additive cancellation in S we see $\varphi(0_R) = 0_S$.
     - Then $0_S = \varphi(0_R) = \varphi(r + (-r)) = \varphi(r) + \varphi(-r)$ so by the uniqueness of additive inverses in S we conclude $\varphi(-r) = -\varphi(r)$.
     - Finally, $\varphi(r_1 - r_2) = \varphi(r_1) + \varphi(-r_2) = \varphi(r_1) - \varphi(r_2)$ by the above calculation.
  3. If $\varphi : R \to S$ is a surjective homomorphism and R has a 1, then S also has a 1 and $\varphi(1_R) = 1_S$. Furthermore, for any unit $u \in R$, the value $\varphi(u)$ is a unit in S whose inverse is $\varphi(u^{-1})$.
     - Proof: Let $s \in S$: then since φ is surjective there exists some $r \in R$ with $\varphi(r) = s$. Then $s\varphi(1_R) = \varphi(r)\varphi(1_R) = \varphi(r1_R) = \varphi(r) = s$, and likewise $\varphi(1_R)s = s$, so $\varphi(1_R)$ is a multiplicative identity in S.
     - For the other part, if u is a unit in R then $1_S = \varphi(1_R) = \varphi(u \cdot u^{-1}) = \varphi(u)\varphi(u^{-1})$, so $\varphi(u)$ is a unit in S with inverse $\varphi(u^{-1})$.
- Associated to a homomorphism are two fundamental objects: the kernel and image.
- Definition: If $\varphi : R \to S$ is a ring homomorphism, the kernel of φ, denoted $\ker\varphi$, is the set of elements in R mapped to 0_S by φ. In other words, $\ker\varphi = \{r \in R : \varphi(r) = 0\}$.
  - Intuitively, the kernel measures how close φ is to being the zero map: if the kernel is large, then φ sends many elements to zero, while if the kernel is small, φ sends fewer elements to zero.
  - Example: The kernel of the reduction homomorphism $\varphi : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ with $\varphi(a) = \overline{a}$ is the subring $m\mathbb{Z}$.
  - Example: The kernel of the evaluation map $\varphi_a : F[x] \to F$ given by $\varphi_a(p) = p(a)$ is the set of polynomials in $F[x]$ with $p(a) = 0$, which is (equivalently) the set of polynomials divisible by $x - a$.
- Definition: If $\varphi : R \to S$ is a ring homomorphism, the image of φ, denoted $\operatorname{im}\varphi$, is the set of elements in S of the form $\varphi(r)$ for some $r \in R$.

- In the context of general functions, the image is often called the range of φ.
- Intuitively, the image measures how close φ is to being surjective: indeed (by definition) φ is surjective if and only if $\operatorname{im}\varphi = S$.

- The kernel and image of a homomorphism are subrings of R and S respectively:
- Proposition (Kernel and Image): Let $\varphi : R \to S$ be a ring homomorphism. Then

1. The image $\operatorname{im}\varphi$ is a subring of S.
   - Proof: Since $\varphi(0_R) = 0_S$, the image contains 0. Furthermore, if s_1 and s_2 are in $\operatorname{im}\varphi$ so that $\varphi(r_1) = s_1$ and $\varphi(r_2) = s_2$ for some $r_1, r_2 \in R$, then $s_1 - s_2 = \varphi(r_1 - r_2)$ and $s_1 s_2 = \varphi(r_1 r_2)$ are also in $\operatorname{im}\varphi$.
   - Thus, $\operatorname{im}\varphi$ contains 0 and is closed under subtraction and multiplication, so it is a subring.
2. The kernel $\ker\varphi$ is an ideal of R.
   - Proof: Since $\varphi(0_R) = 0_S$, the kernel contains 0. Furthermore, if r_1 and r_2 are in $\ker\varphi$ then $\varphi(r_1 - r_2) = \varphi(r_1) - \varphi(r_2) = 0$ and $\varphi(r_1 r_2) = \varphi(r_1)\varphi(r_2) = 0 \cdot 0 = 0$
   - Thus, $\ker\varphi$ contains 0 and is closed under subtraction and multiplication, so it is a subring.
   - Moreover, if $x \in \ker\varphi$ then $\varphi(rx) = \varphi(r)\varphi(x) = \varphi(r)0 = 0$ and likewise $\varphi(xr) = \varphi(x)\varphi(r) = 0\varphi(r) = 0$, so it is an ideal.
3. The kernel is zero (i.e., $\ker\varphi = \{0\}$) if and only if φ is injective. In particular, φ is an isomorphism if and only if $\ker\varphi = \{0\}$ and $\operatorname{im}\varphi = S$.
   - Proof: If $\varphi(a) = \varphi(b)$, then $\varphi(a - b) = \varphi(a) - \varphi(b) = 0$, so $a - b \in \ker\varphi$. Thus, if the only element in $\ker\varphi$ is 0, then we must have $a - b = 0$ so that $a = b$.
   - Conversely, if $x \in \ker\varphi$ and φ is injective, then $\varphi(x) = 0 = \varphi(0)$ implies $x = 0$.
   - The second statement follows from the facts that $\ker\varphi = \{0\}$ is equivalent to φ being injective and $\operatorname{im}\varphi = S$ is equivalent to φ being surjective.

1.3.8 Ideals and Homomorphisms

- Although homomorphisms and quotient rings may not immediately appear to be connected, in fact they are quite deeply related.
  - To begin, observe that if $\varphi : R \to S$ is a ring homomorphism, then the kernel of φ is an ideal of R. Thus, we can use homomorphisms to construct new ideals.
  - Equally importantly, we can also do the reverse: we can use ideals to construct homomorphisms.
  - The key observation in this direction is that the map $\varphi : R \to R/I$ associating a ring element to its residue class (i.e., with $\varphi(a) = \overline{a}$) is a ring homomorphism.
  - Indeed, the two parts of the definition of homomorphism were precisely the properties we arranged for the residue classes modulo I to possess: $\varphi(a + b) = \overline{a + b} = \overline{a} + \overline{b} = \varphi(a) + \varphi(b)$ and $\varphi(a \cdot b) = \overline{a \cdot b} = \overline{a} \cdot \overline{b} = \varphi(a) \cdot \varphi(b)$.
  - Furthermore, the kernel of this map φ is, by definition, the set of elements in R with $\varphi(r) = \overline{0}$, which is to say, the set of elements $r \in I$.
  - Thus, we see that kernels of homomorphisms and ideals are precisely the same things!

- Let us summarize these observations:
- Proposition (Projection Homomorphisms): If I is an ideal of R, then the map $\varphi : R \to R/I$ defined by $\varphi(a) = \overline{a} = a + I$ is a surjective ring homomorphism called the projection homomorphism from R to R/I.
  - Proof: We have $\varphi(a + b) = \overline{a + b} = \overline{a} + \overline{b} = \varphi(a) + \varphi(b)$ and $\varphi(a \cdot b) = \overline{a \cdot b} = \overline{a} \cdot \overline{b} = \varphi(a) \cdot \varphi(b)$, so φ is a homomorphism.
  - Furthermore, φ is surjective, essentially by definition: any residue class in R/I is of the form $\overline{a}$ for some $a \in R$, and then $\varphi(a) = \overline{a}$.

- The next natural question to ask is: if $\varphi : R \to S$ is a homomorphism with kernel I, what can we say about the structure of R/I?
  - For example, if $R = \mathbb{Q}[x]$ and $\varphi : R \to \mathbb{R}$ is defined by $\varphi(p) = p(0)$, then it is easy to see that φ is a homomorphism.
  - Furthermore, the kernel of φ is the ideal I of $\mathbb{Q}[x]$ consisting of the polynomials divisible by x, while the image of φ is the set of rational numbers.
  - Then it is easy to see (from our description of the kernel) that R/I is precisely the same as R/xR, and from the division algorithm for polynomials we know that the residue classes are represented by the polynomials of degree 0 in $\mathbb{Q}[x]$; namely, the constant polynomials $\overline{c}$ for $c \in \mathbb{Q}$.
  - But now notice that the structure of R/I (namely, of $\mathbb{Q}$) is exactly the same as the structure as the image of φ. More formally, these two rings are isomorphic, with an isomorphism given by identifying a residue class $\overline{c}$ with the rational number c.
  - This relabeling can, equivalently, be thought of as being done via the homomorphism φ: we associate the residue class $\overline{c}$ in R/I with the rational number $\varphi(\overline{c}) = c$.
  - In other words: φ gives an isomorphism between $R/\ker\varphi$ and the image $\operatorname{im}\varphi$.
- Theorem (First Isomorphism Theorem): If $\varphi : R \to S$ is a homomorphism of rings, then $R/\ker\varphi$ is isomorphic to $\operatorname{im}\varphi$.
  - Intuitively, φ is a surjective homomorphism $\varphi : R \to \operatorname{im}\varphi$. To turn it into an isomorphism, we must "collapse" its kernel to a single element: this is precisely what the quotient ring $R/\ker\varphi$ represents.
  - Proof: Let $I = \ker\varphi$. We use φ to construct a map $\psi : R/I \to \operatorname{im}\varphi$, and then show that it is injective and surjective.
  - The map is defined as follows: for any residue class $\overline{r} \in R/I$, we define $\psi(\overline{r}) = \varphi(r)$.
  - We must verify that this map ψ is well-defined, so suppose that r' is some other representative of the residue class $\overline{r}$: then $r' - r \in I$, so $\varphi(r' - r) = 0$ and thus $\varphi(r') = \varphi(r)$.
  - Thus, $\psi(\overline{r'}) = \varphi(r') = \varphi(r) = \psi(\overline{r})$, so the map ψ is well-defined.
  - It is then easy to see ψ is a homomorphism, since $\psi(\overline{r} + \overline{s}) = \varphi(r + s) = \varphi(r) + \varphi(s) = \psi(\overline{r}) + \psi(\overline{s})$ and likewise $\psi(\overline{r} \cdot \overline{s}) = \varphi(r \cdot s) = \varphi(r) \cdot \varphi(s) = \psi(\overline{r}) \cdot \psi(\overline{s})$.
  - Next, we see that $\psi(\overline{r}) = 0$ precisely when $\varphi(r) = 0$, which is to say $r \in \ker(\varphi)$, so that $\overline{r} = \overline{0}$. Thus, the only element in $\ker\psi$ is $\overline{0}$, so ψ is injective.
  - Finally, if s is any element of $\operatorname{im}\varphi$, then by definition there is some $r \in R$ with $\varphi(r) = s$: then $\psi(\overline{r}) = s$, meaning that ψ is surjective.
  - Since ψ is a homomorphism that is both injective and surjective, it is an isomorphism.
- By using the first isomorphism theorem, we can construct isomorphisms of rings.
  - In order to show that R/I is isomorphic to a ring S, we search for a surjective homomorphism $\varphi : R \to S$ whose kernel is I.
- Example: If R is any commutative ring, show that $R[x]/(x)$ is isomorphic to R.
  - Let $\varphi : R[x] \to R$ be the "evaluation at 0" homomorphism $\varphi(p) = p(0)$. This map is clearly surjective since for any $r \in R$ we have $\varphi(r) = r$.
  - Furthermore, the kernel of this homomorphism is precisely the collection of polynomials $p(x) = a_0 + a_1x + \cdots + a_nx^n$ with $p(0) = 0$, which is easily seen to be the ideal $I = (x)$ consisting of polynomials divisible by x.
  - Thus, by the first isomorphism theorem, for $I = (x)$ we have $R[x]/I \cong R$.
- Example: Show that $\mathbb{Z}/12\mathbb{Z}$ is isomorphic to $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$.
  - We seek a surjective homomorphism $\varphi : \mathbb{Z} \to (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$ whose kernel is $12\mathbb{Z}$.

- Once this idea is suggested, it is not hard to come up with a candidate, namely, $\varphi(a) = (a \bmod 3, a \bmod 4)$.
- It is easy to verify that map is a homomorphism (since the individual maps of reduction mod 3 and reduction mod 4 are homomorphisms) and it is likewise fairly easy to see that the map is surjective by checking that the images of 0, 1, ... , 11 represent all of the elements in $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$.
- Finally, the kernel of the map consists of all integers a with $\varphi(a) = (0, 0)$, which is equivalent to saying $a \equiv 0 \pmod 3$ and $a \equiv 0 \pmod 4$, so that $3|a$ and $4|a$: thus, the kernel is precisely $12\mathbb{Z}$.
- Therefore, by the first isomorphism theorem applied to this map φ, we conclude that $\mathbb{Z}/12\mathbb{Z}$ is isomorphic to $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$.
- Remark: In fact, we could have avoided checking surjectivity explicitly by instead observing that the first isomorphism theorem yields an injective homomorphism $\psi : \mathbb{Z}/12\mathbb{Z} \to (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$, which must therefore also be surjective since there are 12 elements in both sets.

- We can use the first isomorphism theorem to establish several other related theorems collectively known as the "isomorphism theorems", which characterize how isomorphisms relate to the various ring operations:

- Theorem (Second Isomorphism Theorem): If A is a subring of R and B is an ideal of R, then $A+B = \{a+b : a \in A, b \in B\}$ is a subring of A, $A \cap B$ is an ideal of A, and $(A + B)/B$ is isomorphic to $A/(A \cap B)$.
  - Proof: Clearly $A + B$ contains 0 and $(a + b) - (a' + b') = (a - a') + (b - b')$ so it is also closed under subtraction. For multiplication, we observe $(a + b)(a' + b') = aa' + ba' + ab' + bb'$: the first term is in A since A is a subring, while the other three terms are in B (hence so is their sum) since B is an ideal.
  - For the last statement, consider the map $\varphi : A \to (A + B)/B$ defined by $\varphi(a) = a + B$. This map is well-defined and a homomorphism by the basic properties of quotient rings, and it is surjective since for any class $r + B$ in $(A + B)/B$ for some $r = a + b \in A + B$, we have $\varphi(a) = a + B = r + B$.
  - The kernel of the map φ consists of all $a \in A$ with $a + B = 0 + B$, which is (by definition) equivalent to saying $a \in B$: thus, $\ker \varphi = A \cap B$. In particular, $A \cap B$ is an ideal since it is a kernel of a homomorphism.
  - Thus, by applying the first isomorphism theorem to φ, we see that the rings $A/(A \cap B)$ and $(A + B)/B$ are isomorphic, as claimed.

- Theorem (Third Isomorphism Theorem): If I and J are ideals of R with $I \subseteq J$, then J/I is an ideal of R/I and $(R/I)/(J/I)$ is isomorphic to R/J.
  - Proof: Define the map $\varphi : R/I \to R/J$ given by setting $\varphi(r + I) = r + J$. This map is well-defined because if $r' + I = r + I$, then since J contains I, we also have $r' + J = r + J$, and it is also surjective since for any class $r + J$ in R/J, we clearly have $\varphi(r + I) = r + J$.
  - Furthermore, φ is a homomorphism by the basic properties of quotient rings, since for example $\varphi((r_1 + r_2) + I) = (r_1 + r_2) + J = (r_1 + J) + (r_2 + J) = \varphi(r_1 + I) + \varphi(r_2 + I)$, which shows that φ is additive because $(r_1 + I) + (r_2 + I) = (r_1 + r_2) + I$.
  - Likewise, since $(r_1 + I)(r_2 + I) = r_1r_2 + I$, we see that $\varphi(r_1r_2 + I) = r_1r_2 + J = (r_1 + J)(r_2 + J) = \varphi(r_1 + I)\varphi(r_2 + I)$ and so φ is multiplicative.
  - The kernel of the map φ consists of all $r + I$ in R/I with the property that $r + J = 0 + J$, which is equivalent to saying $r \in J$: thus, $\ker \varphi$ consists of the classes of the form $r + I$ for $r \in J$; this is simply another way of saying that $\ker \varphi = J/I$.
  - Finally, by applying the first isomorphism theorem to φ, we see that the rings $(R/I)/(J/I)$ and R/J are isomorphic, as claimed.

- Example: Inside $R = \mathbb{Z}[x]$, let I be the ideal of all polynomials with zero constant term and J be the ideal of all polynomials with even constant term. Verify the third isomorphism theorem for R, I, and J.
  - As we have already mentioned, both I and J are ideals of R, and clearly $I \subseteq J$.
  - Furthermore, R/I is isomorphic to $\mathbb{Z}$ (per the division algorithm), and J/I is isomorphic to $2\mathbb{Z}$ (the residue classes are represented by the even integers). Also, R/J is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ (since the residue classes are $\overline{0}$ and $\overline{1}$).
  - Then indeed $(R/I)/(J/I) \cong \mathbb{Z}/2\mathbb{Z} \cong R/J$, as claimed.

- Theorem (Fourth/Lattice Isomorphism Theorem): If I is an ideal of R, then there is an inclusion-preserving bijection between subrings A of R containing I and the subrings $\overline{A} = A/I$ of R/I. Furthermore, a subring A of R containing I is an ideal of R if and only if A/I is an ideal of R/I.
    - Proof: We showed during the proof of the second isomorphism theorem that if A contains I then I is an ideal of A, so the association of A with $\overline{A} = A/I$ is well-defined. Conversely, if S is a subring of R/I, then $A = \{r \in R : r + I \in S\}$ is the unique subring of R containing I with the property that $A/I = S$.
    - Furthermore, if B is a subring containing A, then $\overline{B} = B + I$ contains $\overline{A} = A + I$, so the association preserves containment.
    - For the statements about ideals, we showed during the proof of the third isomorphism theorem that if J is an ideal containing I then J/I is an ideal of R/I. Conversely, if J/I is an ideal of R/I, then for any $r \in R$ and $x \in J$ we have $r(x+I) \in J/I$, and this is equivalent to saying that $rx \in J$: thus, J is an ideal of R (since it is already a subring, per the above).
- Example: For $R = \mathbb{Z}$ and $I = 10\mathbb{Z}$, identify the ideals of R containing I and verify they all yield ideals of R/I.
    - The ideals of R containing I are $\mathbb{Z}$, $2\mathbb{Z}$, $5\mathbb{Z}$, and $10\mathbb{Z}$.
    - The corresponding ideals of $R/I = \mathbb{Z}/10\mathbb{Z}$ are $\mathbb{Z}/10\mathbb{Z}$, $2\mathbb{Z}/10\mathbb{Z} = \{\overline{0}, \overline{2}, \overline{4}, \overline{6}, \overline{8}\}$, $5\mathbb{Z}/10\mathbb{Z} = \{\overline{0}, \overline{5}\}$, and $10\mathbb{Z}/10\mathbb{Z} = \{\overline{0}\}$. As claimed, each of these is indeed an ideal of $\mathbb{Z}/10\mathbb{Z}$.

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