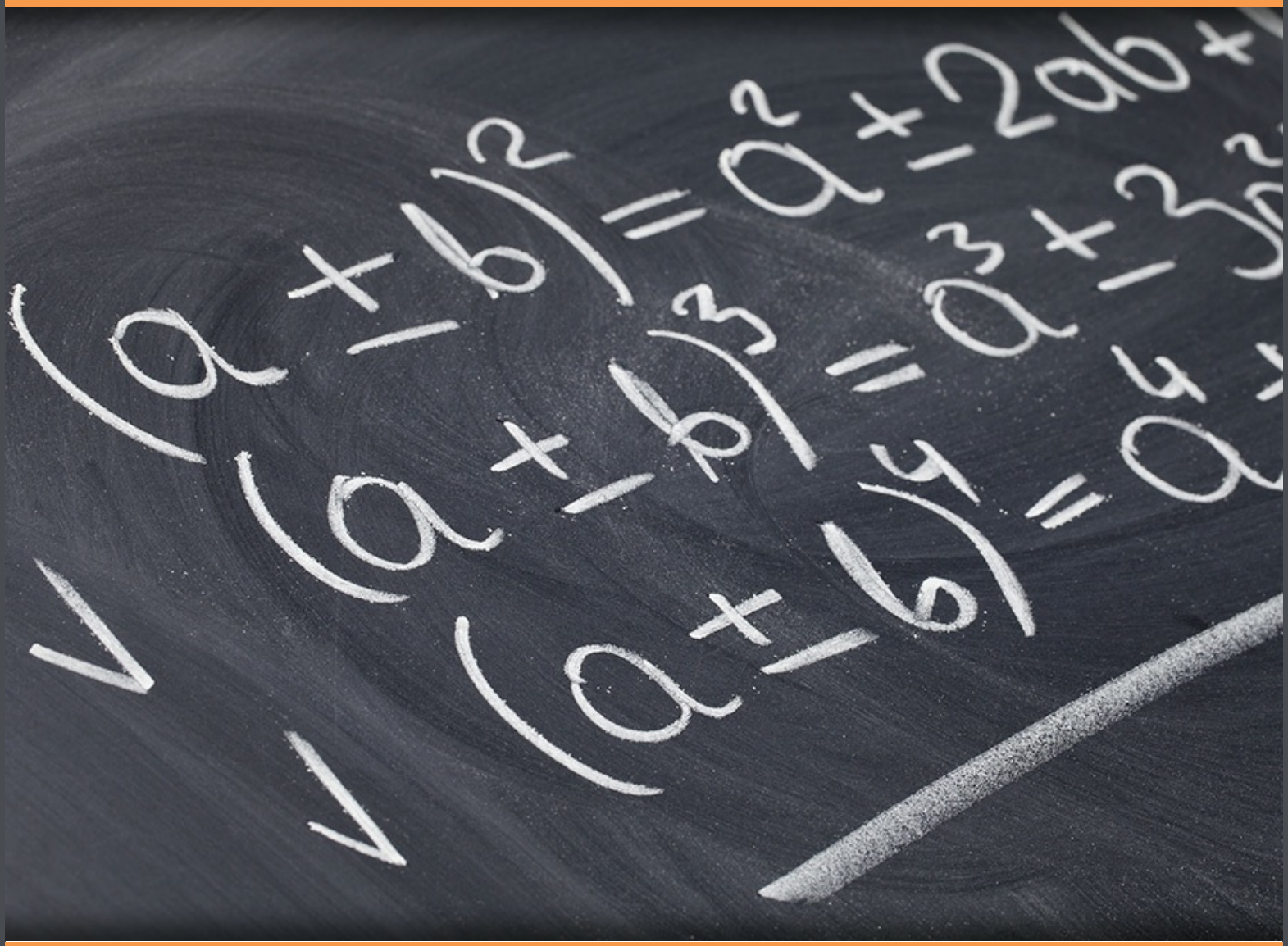
An Introduction to Abstract Algebra

C.K. Taylor



C.K. Taylor

An Introduction to Abstract Algebra

An Introduction to Abstract Algebra
First edition

ISBN 978-87-403-0367-4

Contents

1 Preliminaries

1.1 Introduction to Abstract Algebra

It's always interesting to hear the reaction after telling people that you are fairly far along in your undergraduate mathematical career, and you're taking an algebra course. Reactions range from shock – "Is there really that much to study in algebra?" – to general approval – "Well algebra was the one thing that I was good at in math and the last thing I understood."

What is probably missing from these individuals' understanding is the word "abstract." The further that one goes into mathematics, the more abstract that things get. The focus becomes on the qualities or characteristics that unify and transcend any specific example or instance. To get an example of the spirit of this sort of thing, we will look at the concept of the addition of numbers.

When you first learn how to count, you most likely used positive whole numbers. Addition was done by physically counting objects. At some point, you expanded your set of numbers. Zero was added to this set, as were fractions. Eventually you found out about negative numbers. But these are not the only numbers out there. A number such as $\sqrt{2}$ or π cannot be written as a fraction. We include these numbers under the title of real numbers. While being able to use these numbers is an improvement, there are other mathematical concepts, such as $\sqrt{-1}$ that are not described by using real numbers alone. So we expand our concept of number yet again to include what are known as complex numbers.

Through this process, our concept of number has been stretched and expanded. What was once something that matched up with the fingers on our hand becomes something that while still useful is not as easy to visualize and represent. During our journey from counting numbers to fractions to real numbers and beyond, we have abstracted the idea of number. In the same way we will abstract our conception of algebra until it becomes something much more foreign to us than $3x + 1 = 5$, solve for x. Just as broadening our understanding of number allows us more flexibility in applications (just think of all of the places that a decimal number showed up today in your life), abstract algebra becomes a very useful tool for a wide variety of applications. A few of these follow.

1.1.1 Roots of Polynomials

One goal of algebra, present at the beginning of the subject, is to solve equations for an unknown quantity. This unknown is typically represented by a variable x. Linear equations, characterized by the highest power of x being the first power, are very straightforward to solve. An example would be $ax + b = c$, where a, b, c are constant values with a not equal to zero. The method of solution is to first subtract b from both sides, giving $ax = b - c$ and then divide both sides by a, leaving us with the solution $x = \dfrac{b-c}{a}$.

Not all algebraic equations are linear. We can have higher powers of our variable. In a quadratic equation the highest power of x with a nonzero coefficient is two. The goal is to solve $ax^2 + bx + c = 0$ for x. The solution here is a little harder to come by, and involves a process known as "completing the square." The idea is that because there is a x^2, we will need to take a square root. But because of the presence of the x in our equation, we need to rewrite our equation with one side as a perfect square. Here are the steps to solving a quadratic equation:

1. Divide both sides by a. This is possible because by definition of a quadratic equation, $a \neq 0$. This gives us $x^2 + \left(\frac{b}{a}\right) x + \left(\frac{c}{a}\right) = 0$
2. Subtract $\frac{c}{a}$ from both sides of the equation and $x^2 + \left(\frac{b}{a}\right) x = -\left(\frac{c}{a}\right)$
3. Add $\left(\frac{b}{2a}\right)^2$ to both sides of the equation. This gives us

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

   This is the step of the process that goees by the title "completing the square." The reason why has to do with the form of the left hand side of the equation. If we were asked to expand $(y + z)^2$ we would have $(y + z)^2 = (y + z)(y + z) = y^2 + yz + zy + z^2 = y^2 + 2yz + z^2$. So any algebraic expression that is in the form $y^2 + 2yz + z^2$ is actually a perfect square.

4. With this in mind, we factor

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

5. We also simplify the right hand side of our equation by obtaining a common denominator for the two fractions:

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

6. Take the square root of both sides:

$$x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}}$$

7. Since $\sqrt{\frac{y}{z}} = \frac{\sqrt{y}}{\sqrt{z}}$ we can simplify the right hand side of the equation:

$$x + \frac{b}{2a} = \pm\frac{\sqrt{b^2 - 4ac}}{2a}$$

8. All that remains is to solve for x:

$$x = -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}$$

and simplify:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

And we have the quadratic formula.

It is probably helpful to recap what we have done. Using only basic algebra of balancing both sides of our equation and taking the square root of both sides of the equation, we can determine the value of x as long as we know three numbers: the values of the constants $a, b,$ and c. The equivalent of the quadratic formula has been known at least since 700 AD. This is not too surprising as there are many real world applications where the solution of the quadratic formula is important.

When it comes to deriving formulas for algebraic equations, the quadratic is where many people stop. But there are other types of equations that are out there to be solved. If we look at a cubic equation of the form $ax^3 + bx^2 + cx + d = 0$, we may ask if the same treatment of the quadratic would produce a solution for x. After a little bit of thought we would find that our previous method of completing the square will no longer work. After all, there is now a cubic term in our equation. The solution for the cubic equation had to wait another 800 years or so, but in 1545, amidst a web of intrigue, Cardano published the solution of the cubic equation. The cubic formula is much more complicated than that of the quadratic formula, however it works in the same way as the quadratic. Both formulas only require us to know the coefficients of our equation. We plug these numbers into a formula that combines basic arithmetic and roots of certain degrees – called radicals, and the formula gives us the value of x.

What about equations where x^4 is the highest power? In the process of finding a method to solve a cubic equation, a similar method was found for quartic equations of the form $ax^4 + bx^3 + cx^2 + dx + e = 0$. The solution to this was also published in 1545.

The question that arises from this is, "Is it possible to use similar methods to solve a quintic equation of the form $ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0$? " This is equivalent to asking, "Is there an equation involving only the coefficients of the quintic equation that produces the value of x? The complete answer to this question had to wait until 1822, when Galois – also no stranger to intrigue – showed that although we can use basic algebra to solve some quintics, in general quintic equation cannot be solved using algebra radicals.

Proving a negative is generally hard, but Galois was able to show that there is no solution of a general quintic using algbraic techniques by employing abstract algebra. An entire subfield of mathematics, called Galois theory, is named after him.

Of course one might expect that abstract algebra can be used to answer questions of an algebraic nature. What is not so obvious is that it can be used to tell us things about other areas of mathematics.

1.1.2 Straight Edge and Compass Constructions

Plane geometry was developed in antiquity by the Greeks. One feature of this geometry is the desire to construct idealized geometric figures by use of two tools, the compass and straightedge with no markings. A compass can be used to draw arcs and circles. An unmarked straightedge can be used to draw lines, but without the ability to measure the length of those lines. With these tools and a few rules in place, the goal was to perform certain geometric constructions.

It is relatively easy to begin with an arbitrary angle and bisect it, or split it into two angles of equal measure. The question that arose from this was, "Is it possible to trisect an arbitrary angle?" In other words, if we are given the angle θ, then is it possible to construct the angle $\theta/3$? While this is possible for certain values of θ, it was unknown if this could be done for an arbitrary angle. We note that the absence of a solution does not mean its nonexistence, only that it has not been discovered yet. In 1837 Wantzel demonstrated that it is in fact impossible to trisect a given angle. What is surprising about this is that the proof of a geometric fact involves the use of abstract algebra.

1.1.3 Other Applications and a Brief Note

Other areas of mathematics heavily depend upon abstract algebra, which is why most graduate programs require students to take several high-level algebra courses. But abstract algebra is found in a multitude of disciplines. Theoretical physicists employ the language of group theory in their models of how the universe works. Symmetries in chemistry can be represented abstractly using the language of abstract algebra. Even the topic of codes employs abstract algebra.

Of course it takes a little bit of study to get to any of these exciting applications. The goal of this book is to bring you to a place where you understand why certain geometric constructions are impossible. Most of what follows in the remainder of this chapter will be a quick review of things that you've probably seen throughout your mathematical career. This material can sometimes seem a little dull, but just because something is uninteresting does not mean that it is unimportant.

There is a systematic building that goes on in abstract algebra. Other definitions and topics build upon the very basic concepts (that manage to trip some people up) and proof strategies of this chapter. We must be certain to have a firm foundation to do any subsequent building. So let's get started!

1.1.4 Exercises

1. Solve the quadratic $3x^2 - 8x + 10$ by completing the square and working through the steps of the derivation of the quadratic formula (don't just plug the coefficients into the quadratic formula).
2. Research the cubic equation and use it to solve $2x^3 + 5x - 7 = 0$
3. Find other real world applications of abstract algebra.
4. Write a brief summary of the life and mathematical contributions of Cardano.
5. Write a brief summary of the life and mathematical contributions of Galois.

1.2 Logic and Proof

"I know what you're thinking about," said Tweedledum; "but it isn't so, nohow." "Contrariwise," continued Tweedledee, "if it was so, it might be; and if it were so, it would be; but as it isn't, it ain't. That's logic."

Through the Looking Glass by Lewis Carroll

At its most fundamental level, mathematics involves statements about certain objects. These objects can be numbers, polygons, or things that are so abstract that they cannot be listed out, drawn, or visualized. From a handful of statements concerning these objects, we attempt to form other statements. The process by which we do this is to use deductive logic. Deductive logic proceeds in an orderly way through statements. A string of these statements forms an argument or proof. Valid proofs (the ones that we are interested in) have a conclusion that follows logically from all of the prior statements or hypotheses.

Unlike other fields of knowledge, a mathematician can prove definitively that he or she is absolutely correct. Provided that the hypotheses are true and the argument form is valid, the conclusion must be true. This form of thought has been with us since ancient Greece, and the fundamental principles of logic laid down by Aristotle are still with us today. The statements concerning numbers, proportions, and the sorts of things we will encounter in this book were never dreamed of in antiquity, but the logic and arguments structures that hold it all together have been part of our cultural history for centuries.

It is assumed that you have seen some sort of logic before. This may have been in a proofs or logic course, or you may have learned it by example of seeing it done in a math class. In this section we will look at the main proof strategies that will be used throughout the course.

1.2.1 Direct Proof

The first proof strategy that we will examine is called a *direct proof*. In this type of proof our goal is to show that the statement "If P then Q" is true. Here P and Q are themselves statements, meaning that they are sentences that can be classified as either true or false. The method of direct proof to prove "If P then Q" involves the steps:

1. Begin by assuming that the statement P is true.
2. Use other information that we know from mathematics to establish that the statement Q is also true.

Example: Use a direct proof to show that for any odd integer n, n^2 is also odd.

Before proceeding with a proof we will formalize our problem. Implicit in this is that we know what an integer is, and what an odd number is. Integers are positive and negative whole numbers. An odd number is of the form $2k+1$ where k is an integer. What the above problem is asking us to do is to prove: if n is odd, then n^2 is odd.

We begin by supposing that n is an odd integer. Thus it has the form $n = 2k+1$ where k is an integer. Now our goal is to show that n^2 is also odd. We do this directly by squaring n and seeing where the algebra leads us:

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$$

We now use some known properties about integers: the product of any two integers is an integer, and the sum of any two integers is an integer. This shows us that $n^2 = 2(2k^2 + 2k) + 1$ is in the form $2M + 1$ where M is an integer, and thus n^2 is an odd number.

□

1.2.2 Indirect Proof

Proving a mathematical statement with a direct proof is not the only method of proof. We may also use one of two indirect methods of proof: proof of the contrapositive or proof by contradiction. We will begin by looking at the contrapositive.

Definition:

The logical statement "If *P* then *Q*" is logically equivalent to its *contrapositive*: "If not *Q* then not *P*".

□

Example: The contrapositive of the statement "If it is raining, then I will take my umbrella to school" is "if I did not take my umbrella to school, then it is not raining."

□

To prove the statement "If *P* then *Q*" by use of the indirect method of proof that uses the contrapositive, we use the following process:

1. From the statement "If *P* then *Q*" form the contrapositive "If not *Q* then not *P*."
2. Assume that "not *Q*" is true and from this use a method of direct proof to demonstrate that "not *P*" is true.

What follows is an example of a contrapositive proof. Note that this involves first forming the contrapositive.

Example: Prove by use of a contrapositive that the following is true: For any integer n, if n^2 is odd then n is odd.

We form the contrapositive of the above statement and obtain "for any integer n, if n is not odd, then n^2 is not odd." We can smooth this out by rephrasing the "not odd" as "even." So in order to prove the original statement, we must show that if n is an even number then n^2 is an even number.

Suppose that n is an even integer. By definition, it is of the form $2k$ where k is an integer. We use basic algebra and see that

$$n^2 = (2k)^2 = 4k^2 = 2(2k^2)$$

Thus $n^2 = 2(2k^2)$ and is an even number (again since the product of any two integers is also an integer). This not only shows that for any integer n "if n is even, then n^2 is even," it also shows that for any integer n "if n^2 is odd then n is odd."

□

WARNING: The statement "If *P* then *Q*" is not logically equivalent to the *converse* "If *Q* then *P*."

□

We now consider proof by contradiction. This is another indirect method of proof, but has a different structure than a contrapositive proof.

For proof by contradiction of the statement "If *P* then *Q*"

1. Begin by assuming that both *P* and not *Q* are true statements.
2. Use other known facts to show that this implies a contradiction.

Most statements that can be proved with a contrapositive proof can also be proved by contradiction.

Example: Prove the following by contradiction: "For any integer n , if n^2 is odd then n is odd."

We begin by supposing that n^2 is an odd integer and n is not odd. In other words, n is even. If n is an even integer, then it is of the form $n = 2k$. We square n and see:

$$n^2 = (2k)^2 = 2(2k^2),$$

which is an even number. We have reached a contradiction, as we simultaneously have that n^2 is odd and n^2 is even. Our original supposition was incorrect, and thus we have proved the statement "if n^2 is odd, then n is odd."

□

Note: We may combine this statement with the statement from the example that we opened the section with:

- "If n is odd, then n^2 is odd."
- "If n^2 is odd, then n is odd."

Basic logic tells us that these two statements are equivalent to saying "n is odd if and only if n^2 is odd." This fact comes into play when we are asked to prove the statement "*P* if and only if *Q*." This really means that we need to prove two statements: "If *P* then *Q*" AND "If *Q* then *P*."

□

For a more sophisticated example of a proof by contradiction, we look at a classical example that can be found in geometry textbook par excellence, Euclid's *Elements*.

Theorem 1. *The set of prime numbers is infinite.*

The proper development of this proof would require a definition of a prime number. In addition to this we need the fact that every number is a prime number or a product of several primes. We note that a natural number is prime if it has exactly two divisors. The set of prime numbers thus includes 2, 3, 5, 7, and 11. There is nothing about these facts as stated that imply that the set of prime numbers is infinite or finite. As we look at our set of natural numbers, there are chunks of consecutive numbers that are all composite.

Proof. Assume by way of contradiction that there are a finite number of primes.

Let $S = \{p_1, p_2, \ldots, p_n\}$ denote the set of all prime numbers. Construct $M = p_1 p_2 \ldots p_n + 1$, i.e. the product of every prime with one added to it.

Since $M > p_i$ for all of the primes in S, $M \notin S$ and M is not prime. Thus M has a prime divisor p, where p is one of the primes in our set S. However, if p divides M and p divides $p_1 p_2 \cdots p_n$, then p divides their difference $M - p_1 p_2 \cdots p_n = 1$. This is a contradiction (because no number divides 1 other than 1) and so our original assumption was false.

1.2.3 Mathematical Induction

Mathematical induction is a proof technique that is helpful to prove statements regarding nearly all of the natural numbers $\mathbb{N}$. Every induction proof has two steps: first show that a statement is true for $n = 1$, this is sometimes called the anchor step; second, show that if the statement is true for a general k, then it must be true for $k + 1$ as well.

The process could be thought of as knocking over dominos. We think of the dominos all arranged in a line. To knock them all over, we can push over the first domino, which will fall and hit the second. This second will fall and knock down the third, and so on. Pushing over the first domino is like the anchor step of our induction. Showing that if our statement is true for k then it is also true for $k + 1$ is akin to the k th domino in our line knocking over the $(k + 1)$th.

More formally we have the definition: **Definition:** For each $n \in \mathbb{N}$, let $P(n)$ be a statement about n. The *principle of mathematical induction* states that if both:

1. $P(1)$ is true.
2. For every $k \in \mathbb{N}$, if $P(k)$ is true,then $P(k + 1)$ is true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

□

As always, it's best to see how this process works by doing some examples.

Example: Show that the sum of the first n natural numbers $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$.

It is clear that induction should be used (not just because this is the section of the book about induction). We know this proof should use mathematical induction because we are asked to prove something involving the first n natural numbers.

For $n = 1$: We need to show that the above formula is valid for $n = 1$.

$\frac{1}{2}1(1 + 1) = 1$. So the anchor has been established.

For $k \Rightarrow k + 1$:

Assume by induction that $1 + 2 + \cdots + k = \frac{1}{2}k(k + 1)$. Since we want to prove a statement concerning $1 + 2 + \cdots + k + (k + 1)$ it would be most helpful to add $k + 1$ to both sides of our equation.

$$\begin{aligned} 1+2+\cdots+k+(k+1) &= \tfrac{1}{2}k(k+1)+(k+1) \\ &= \tfrac{1}{2}k(k+1)+\tfrac{1}{2}2(k+1) \\ &= \tfrac{1}{2}[k(k+1)+2(k+1)] \\ &= \tfrac{1}{2}[k^2+k+2k+2] \\ &= \tfrac{1}{2}[k^2+3k+2] \\ &= \tfrac{1}{2}(k+1)(k+2) \end{aligned}$$

Thus the formula holds for $k+1$ and by induction, we have proved the statement for all natural numbers n.

Example: Show that $\sum_{j=1}^{n} 2^j = 2^{n+1} - 2$.

By induction:

For $n = 1$*:*

$\sum_{j=1}^{1} 2^j = 2^1 = 2 = 4 - 2 = 2^{1+1} - 2$. So the anchor has been established.

For $k \Rightarrow k+1$*:*

Assume that $\sum_{j=1}^{k} 2^j = 2^{k+1} - 2$. Now add 2^{k+1} to both sides of the equation. The right side becomes

$$2^{k+1} - 2 + 2^{k+1} = 2 \cdot 2^{k+1} - 2 = 2^{k+2} - 2.$$

Thus the formula holds for $k+1$. By induction, this proves the statement for all n.

Example: Form a conjecture regarding a formula for the sum of the first n odd numbers, and prove that your formula is true.

This problem requires an extra bit of work, as we are not given an explicit formula. To figure out what the formula should be, we need to play a bit. Since we›re trying to form and prove a statement regarding the sum of odd numbers, let›s start by doing a few addition problems.

- The sum of the first odd number is 1.
- The sum of the first two odd numbers is $1 + 3 = 4$.

- The sum of the first three odd numbers is $1 + 3 + 5 = 9$.
- The sum of the first four odd numbers is $1 + 3 + 5 + 7 = 16$.

We could keep doing this, but there is probably enough evidence now to form a guess as to the sum of the first n odd numbers. We see that all of these sums are perfect squares. Our conjecture is: "The sum of the first n odd numbers is n^2." We could also write this as "$1 + 3 + 5 + \cdots + (2n - 1) = n^2$."

Now it's time to prove this conjecture. Since we are dealing with a statement about the natural numbers, we will use mathematical induction. Our work above in formulating the conjecture also serves as an anchor for the induction. We actually have more than we typically do, as we have demonstrated that the statement holds not only for $n = 1$, but also for $n = 2, 3,$ and 4. Of course we are trying to prove a statement about all of the natural numbers, so we have an infinite number of these left to try. That is why we need to do the next part of our inductive proof.

Now we suppose that the sum of the first k odd numbers is k^2. The sum of the first $k + 1$ odd numbers is $1 + 3 + \cdots + (2k - 1) + (2k + 1) = k^2 + (2k + 1)$ by use of our inductive hypothesis. We then use basic factoring and see that $k^2 + 2k + 1 = (k + 1)^2$. By induction we have shown that for any $n \geq 1$ the sum of the first n odd numbers is n^2.

□

Note: Notice that there is nothing special about starting at $n = 1$. Induction could be anchored for a higher initial value k_0 of n, then we could proceed as normal. The end result would be a true statement for all $n \geq k_0$

□

Example: Show that for all natural numbers $n \geq 4, \quad 2^n < n!$

We begin by noting that the above inequality is not true for $n = 1, 2, 3$. We must start the proof by anchoring at $n = 4$:

For $n = 4$:

$2^4 = 16 < 24 = 4!$

For $k \Rightarrow k + 1$:

We assume that $2^k < k!$. Multiply both sides of the inequality by 2 and obtain $2 \cdot 2^k < 2 \cdot k!$. Now $2 \cdot 2^k = 2^{k+1}$. Furthermore, since *k* is a natural number, 2 ≤ (*k* +1) and so 2 · *k*! ≤ (*k* +1) · *k*!. Thus:

$$2^{k+1} = 2 \cdot 2^k < 2 \cdot k! \leq (k + 1) \cdot k! = (k + 1)!$$

We have shown that $2^{k+1} < (k + 1)!$. Since the statement holds for $k + 1$, by induction it is true for all $n \geq 4$.

□

1.2.4 Exercises

1. Assume that the only prime numbers that you knew were $\{2, 3, 5, 7\}$. Work through Euclid's proof of the infinitude of primes by assuming this set is the set of all of the prime numbers. What contradiction do you arrive at?
2. Prove that the integer n is divisible by 5 if and only ifn^2 is divisible by 5.
3. Without the help of a calculator or computer, prove that the number

   $$123451234512345^3 - 123451234512345$$

   is divisible by 6. [HINT: There is nothing special about the number 123451234512345. The problem could have asked to demonstrate that $n^3 - n$ for any integer n.]

4. Prove that for all natural numbers n: $1 + 2 + 3 + \cdots + (n - 1) + n = \frac{n(n+1)}{2}$

5. Produce a formula in terms of n for the sum of the first n even numbers. Use mathematical induction to prove your formula is correct. [Hint: To arrive at your formula, you may want to use the previous exercise]
6. Prove that for all natural numbers n : $1 + 4 + 9 + \cdots + (n-1)^2 + n^2 = \frac{n(n+1)(2n+1)}{6}$
7. Produce a formula in terms of n for the sum $1 + 7 + 19 + \cdots + (3n^2 - 3n + 1)$ and prove that your formula is correct by use of mathematical induction.
8. Prove by mathematical induction that for all natural numbers n :
$$(x_1 + x_2 + \cdots x_n) \leq (x_1^n + x_2^n + \cdots x_n^n)^{1/n}$$
where x_i are all real numbers.

1.3 Set Theory

If you were stranded on a desert island with a friend and wanted to pass the time you might think of trying to play chess. But since there probably isn't a chess set on the island, you would have to improvise. It would be easy enough to draw a board in the sand. Rocks could be used for bishops, a coconut for the kings, and so on. It wouldn't matter that what you were using didn't match a traditional chess set. What would be important is that you and your friend would have an understanding of how each item represented a particular piece on a chessboard. A rock would only move diagonally, like a bishop on a traditional chessboard. The definition of what it is to be a bishop would be of the utmost importance. If you recorded your moves and were eventually rescued, another chess aficionado safe at home in his study could follow the movements of rocks and coconuts in the sand by knowing the sequence of moves that you made.

What does chess have to do with abstract algebra? In the above story each object has a well-defined role in the game of chess. It is not important that a rook looks like a castle, only that what we use as a rook moves on our board in the sand the same way that a rook moves on a traditional chessboard. The concept of well-defined ideas is very important throughout all of mathematics. In any field of mathematics, it is imperative that we are all working with the same set of concepts and definitions. This is another one of the features that sets mathematics apart from other fields of knowledge. Ideas can be expressed without any ambiguity whatsoever.

In addition to the use of logic, mathematics is built upon the language of set theory. A good grasp of this area of math is important for the study of any other areas.

1.3.1 Sets

In mathematics a set is a well-defined collection of objects, which are known as elements. These elements can be anything – numbers, letters, or even other sets. What is crucial is that we can unambigously determine what elements are in the set, and what elements are not in the set.

While a course such as Calculus involves sets of real numbers, in abstract algebra our sets are in some ways more basic. Most of what we consider initially will be sets with a finite number of elements.

In crafting proofs and arguments, it is helpful to have some notation to serve as shorthand. You have already seen this throughout your mathematical career. Rather than writing "x is greater than 5," the greater than symbol can be used to write "$x > 5$." In a similar way we have the following notation regarding sets.

Notation:

$x \in A$	"x is an element of the set A"	
$x \notin A$	"x is not an element of the set A"	
$A \subseteq B$	"A is a subset of B"	every element in A is an element of B.
$A \subset B$	"A is a proper subset of B"	If $x \in A$ then $x \in B$, and there is at least one element $x \in B$ for which $x \notin A$.
$A = B$	"A is equal to B"	A and B contain the same elements.

□

Note: To show that two sets, *A* and *B*, are equal to each other, we must show $A \subseteq B$ and $B \subseteq A$.

□

Example: Let $A = \{x \in \mathbb{Z} \mid x^2 - 3x + 2 = 0\}$ and let $B = \{1, 2\}$. Prove that $A = B$.

We begin by showing that $B \subseteq A$. We note that $(1)^2 - 3(1) + 2 = 0$ and $(2)^2 - 3(2) + 2 = 0$, so $\{1, 2\} \subseteq A$. To show that $A \subseteq B$, we suppose by way of contradiction that $A \nsubseteq B$. That is, there is an element $y \in A$ and $y \notin B$. If $y \in A$, by definition $y^2 - 3y + 2 = 0 \Rightarrow (y - 2)(y - 1) = 0$, and so $y = 1$ or $y = 2$. In either of these cases $y \in B$, and so we have a contradiction.

We have that $A \subseteq B$ and that $B \subseteq A$. This double inclusion demonstrates that $A = B$.

□

The following are abbreviations for sets that will be used throughout the book. They are more or less standard across mathematics:

Notation:

$\mathbb{N}$	$:= \{1, 2, 3, \cdots\}$	The Natural Numbers
$\mathbb{Z}$	$:= \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$	The Integers
$\mathbb{Q}$	$:= \{p/q : p, q \in \mathbb{Z} \text{ and } q \neq 0\}$	The Rational Numbers
$\mathbb{R}$		The Real Numbers
$\mathbb{C}$	$:= \{a + bi \vert a, b \in \mathbb{R}, i = \sqrt{-1}\}$	The Complex Numbers

□

1.3.2 Set Operations

The study of arithmetic involves the basic operations of addition, subtraction, multiplication, and division. For each of these operations, we begin with two numbers, apply the operation, and this gives us a number as a result. In a similar way we can begin with two sets, apply a set operation, and this gives us another set. Set theory lies at a deeper level than arithmetic, and it is even possible to define our arithmetic in terms of set theory operations.

We begin with a *universal set*. Just as the universe is the totality of the physical world, the universal set for a particular problem is the set of all elements that we can choose from to form other sets. There is not one universal set. The universal set that we use depends upon the context of our problem.

Example: Let A be the set of numbers such that $x^2 = 16$.

Here the set A is very much dependent upon universal set that we use. If the universal set is the set of positive whole numbers, then $A = \{4\}$. If the universal set is the set of positive and negative whole numbers, then $A = \{-4, 4\}$.

□

We will now look at set operations, and the process of forming new sets from other ones.

Definition: For a given universal set *X* and two sets $A \subseteq X$, $B \subseteq X$

- The *union* of sets *A* and *B* is $A \cup B := \{x : x \in A \text{ or } x \in B\}$
- The *intersection* of sets *A* and *B* is $A \cap B := \{x : x \in A \text{ and } x \in B\}$
- The *complement of B relative to* A is $A \backslash B := \{x : x \in A \text{ and } x \notin B\}$
- The *complement of A* is $A^C := X \backslash A = \{x : x \in X \text{ and } x \notin A\}$

Note: The word *or* has a couple of different uses in the English language. In the exclusive sense, it can imply a choice between two options. In the inclusive sense, it means that you can choose between either of the options or both. For example of both of these senses of the word*or*, suppose you are at dinner. If a waiter asks you if you want a chicken or beef, the implication is that you can order one of these items. On the other hand if you are asked if you want butter or sour cream on your baked potato, the assumption is that you can have either of these items or both. Obviously it will not do to carry over this ambiguity in our mathematical language. In mathematics, unless specifically told otherwise, the word or is used in the inclusive sense. Thus if $x \in A \cup B$, then x can be an element *A*, an element of *B*, or of both *A* and *B*.

□

Example: Let $A = \{1, 2, 3, 4, 5\}, B = \{1, 3, 5, 7\}$ with universal set $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$

- $A \cup B = \{1, 2, 3, 4, 5, 7\}$
- $A \cap B = \{1, 3, 5\}$
- $A \backslash B = \{2, 4\}$
- $B \backslash A = \{7\}$
- $A^C = \{6, 7, 8\}$
- $B^C = \{2, 4, 6, 8\}$
- $A \cup X = X$
- $A \cap X = A$
- $X^C = \{\ \}$.

□

The very last item in the above list, a set with no elements, has a variety of properties that are revealed by a little bit of thought.

There are many properties of the empty set revealed by a little bit of thought. For any set *A* and universal set *X*:

- $\{\ \}^C = X$
- $\{\ \} \cup A = A$
- $\{\ \} \cap A = \{\ \}$

As the following theorem will show, it is appropriate to talk about *the* empty set and not *an* empty set.

Theorem 2. *The empty set is unique.*

As this is our first uniqueness proof, we should draw attention to the proof strategy we will use. Any time that we want to show something is unique or one of a kind, it is typically a good idea to use a proof by contradiction. That is, we will assume that something is not unique (there are at least two of them) and then arrive at a contradiction.

Proof. Assume by way of contradiction that there are two empty sets E, F where $E \neq F$. We look at the set $E \cup F$. Since E is empty, $E \cup F = F$. However, since F is empty $E \cup F = E$. Thus we have $E = E \cup F = F$. This is a contradiction so our original assumption was false.

□

Definition: The *empty set* is the set with no elements in it. It is denoted $\emptyset$

□

One last property of the empty set, that takes slightly more thought is that for any set A, $\emptyset \subseteq A$. Why is this true? Well one and only one of the following are true:

1. For all sets A, $\emptyset \subseteq A$
2. There is a set A such that $\emptyset \not\subseteq A$.

If #2 is true, then since $\emptyset \not\subseteq A$ there is an element x such that $x \in \emptyset$ and $x \notin A$. By definition of the empty set, there can be no element $x \in \emptyset$. Thus #2 is false, and #1 must be true.

The is an example of a statement that is vacuously true. It is similar to the situation of a man who tries to impress his date by telling her, "All the Ferraris in my garage are red." The only way that this statement is false is if he has a Ferrari in his garage that is not red. The statement is true if he does indeed have a red Ferrari in his garage. It is also true if does not have a Ferrari in his garage at all.

Another topic in set theory that is worth mentioning are De Morgan's Laws. De Morgan's Laws are two statements pertaining to how the union, intersection, and relative complement interact with one another. They show up in a number of places. We are interested in them here mainly so that we can practice using set notation to prove statements.

Theorem 3 (DeMorgan's Laws). For any sets A, B, C:

1. $A\backslash(B \cup C) = (A\backslash B) \cap (A\backslash C)$
2. $A\backslash(B \cap C) = (A\backslash B) \cup (A\backslash C)$

Proof. We will prove #1. To prove that the two sets are equal we must show that $A\backslash(B \cup C)$ and $(A\backslash B) \cap (A\backslash C)$ are subsets of one another.

$$
\begin{aligned}
&\text{Let } x \in A\backslash(B \cup C) && \Rightarrow x \in A \text{ and } x \notin B \cup C.\\
& && \Rightarrow x \in A \text{ and } (x \notin B \text{ and } x \notin C).\\
& && \Rightarrow (x \in A \text{ and } x \notin B) \text{ and } (x \in A \text{ and } x \notin C).\\
& && \Rightarrow (x \in A\backslash B) \text{ and } (x \in A\backslash C).\\
& && \Rightarrow x \in (A\backslash B) \cap (A\backslash C).
\end{aligned}
$$

Thus $A\backslash(B \cup C) \subseteq (A\backslash B) \cap (A\backslash C)$. Now we will show the other inclusion.

If $x \in (A\backslash B) \cap (A\backslash C)$ then $x \in A\backslash B$ and $x \in A\backslash C$. Thus $x \in A, x \notin B$ and $x \in A, x \notin C$. In other words, $x \in A$ and $x \notin B \cup C$. Therefore $x \in A\backslash(B \cup C)$ and $(A\backslash B) \cap (A\backslash C) \subseteq A\backslash(B \cup C)$.

Since we have shown both inclusions, we have proved that the sets are equal. □

The last topic in set theory that will be used in what follows is the Cartesian product.

Definition: Let *S* and *T* be sets. The *Cartesian product* of *S* and *T*, denoted $S \times T$ is the set of all ordered pairs (s,t) where $s \in S$ and $t \in T$.

□

The Cartesian product is used in the Cartesian or rectangular coordinate system when we plot points (x,y) in the plane $\mathbb{R} \times \mathbb{R}$. We will be more interested in using the Cartesian product for some careful definitions as well as constructing some specific examples later on.

Example: Let $S = \{a,b,c\}$ and $T = \{2,3\}$. List all elements of the Cartesian product $S \times T$.

We must form all possible pairs (s,t) where the first element is from the set *S* and the second element is from the set *T*. There are 3 choices for the first element and 2 for the second. So there are $2 \times 3 = 6$ elements in $S \times T$.

$S \times T = \{(a,2),(a,3),(b,2),(b,3),(c,2),(c,3)\}$.

□

Note: The if S is a finite set with m elements and *T* is a finite set with n elements, then $S \times T$ is a finite set with $m \cdot n$ elements. If either *S* or *T* is an infinite set, then $S \times T$ is also infinite.

1.3.3 Exercises

1. Prove that $A \backslash B = A$ if and only if $A \cap B = \emptyset$
2. Prove that for any sets A, B, C: $A \backslash (B \cap C) = (A \backslash B) \cup (A \backslash C)$ [HINT: This is one of De Morgan's Laws]
3. The symmetric difference of the sets A and B is defined as
$$A \Delta B = (A \backslash B) \cup (B \backslash A).$$
Prove that $A \Delta B = (A \cup B) \backslash (A \cap B)$
4. Prove that $(S \cup T) \times (V \cup W) = (S \times V) \cup (S \times W) \cup (T \times V) \cup (T \times W)$

1.4 Mappings and Equivalence Relations

This section is linked by the common theme of examining particular subsets of the Cartesian product $S \times T$. Any subset of $S \times T$ is a relation.

Definition: A *relation* R between the sets *S* and *T* is any subset of the Cartesian product $S \times T$. If $(s,t) \in R$ we say that "*s* is related to *t*" and write *s R t*.

□

We are primarily interested in two types of relations:

- Mappings, which are a generalization of the functions encountered in Calculus.
- Equivalence relations, which are a sort of generalization of equality.

Both of these topics are properly defined in terms of the Cartesian product.

1.4.1 Mappings

One can't go too far into any part of mathematics without bumping into a mapping. Sometimes these mappings go by different, more specialized names. For instance, Calculus is really the study of mappings known as continuous real-valued functions. The functions can be polynomial, trigonometric, logarithmic, and even more complicated than these.

There are a number of ways to intuitively grasp the concept of a real-valued function. One that is helpful is to think of a function as a machine. For every allowable real number that is entered into a real-valued function, there is exactly one real number as an output. We make the qualification that the input must be allowable since there are some real-valued functions for which certain inputs result in an undefined output. For an example of this, try plugging $x = 0$ into $1/x$ and state the number that you end up with.

Now there is nothing that would require us to use subsets of the real numbers as inputs and outputs of our function machine. While some sort of set of numbers is a how functions got their start, there is no reason that we need to restrict ourselves to just using numbers. Mappings generalize our idea of a real-valued function, allowing for *any* sets for input and output. We could even have an input set of, say quadrilaterals, and an output set of numbers. The key is that each allowable input for our mapping may only have one output.

We will carefully define a mapping in this section, as well as look at different specialized features of these mappings. Our study of abstract algebra will require us to examine even more specialized mappings, but we must first understand the basic concepts.

Definition: A *mapping f* from the set S to the set T, denoted by $f : S \to T$, is a subset M of the Cartesian product $S \times T$ where for every $s \in S$ there is exactly one $t \in T$ such that $(s, t) \in M$.

If $(s, t) \in M$ we write $t = f(s)$.

□

Note: This definition allows us to have a well-defined notion of a mapping, but it deemphasizes the intuitive description of mapping as a rule/transformation/machine that assigns a t to each s. Contained in this definition is the fact that every $s \in S$ is paired with one (and only one) $t \in T$.

□

Example: The function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = x^2$ consists of the points

$$M = \{\ldots, (-2, 4), (-1, 1), (0, 0), (1, 1), (2, 4), (3, 9), \ldots\}.$$

□

Definition: Given a mapping $f : S \to T$:

- The set S is the *domain* of the mapping.
- The set T is the *codomain* of the mapping.
- The set $R(f) := \{t \in T | f(s) = t \text{ for some } s \in S\}$ is the *range* of the mapping. By definition $R(f) \subseteq T$.

□

Definition: Given a mapping $f : S \to T$ where for every $x_1, x_2, \in S$ if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$, we say that f is *one-to-one* or *injective*.

□

Note: We may form the contrapositive of the above statement and see that a mapping is one-to-one when $f(x_1) = f(x_2)$ implies that $x_1 = x_2$.

□

Definition: Given a mapping $f : S \to T$ where for every $y \in T$ there exists a $x \in S$ such that $f(x) = y$, we say that f is *onto* or*surjective*.

□

Note: An alternate definition of a surjective mapping $f: S \to T$ is a mapping for which $S = R(f)$ the range of f.

□

Definition: A mapping is *bijective* if it is both injective and surjective.

□

Example:

1. The mapping $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is neither injective nor surjective.

   $f(x) = f(-x)$, but for all nonzero x, $-x \neq x$, so the mapping is not one-to-one.

   The set $a = \{x : x < 0\} \subset \mathbb{R}$ is part of the image of f, so it is not onto.

2. The mapping $f : A \to B$ defined by $f(x) = \frac{1}{\sqrt{x+1}}$ where $A = \{x \in \mathbb{R}\{x \in: x > -1\}$ and $B = \{y \in \mathbb{R}: x > 0\}$ is injective.

   Suppose $f(x_1) = f(x_2)$

   $$\Rightarrow \frac{1}{\sqrt{x_1 + 1}} = \frac{1}{\sqrt{x_2 + 1}} \Rightarrow \sqrt{x_1 + 1} = \sqrt{x_2 + 1} \Rightarrow x_1 + 1 = x_2 + 1 \Rightarrow x_1 = x_2.$$

   Therefore f is injective.

3. The mapping $f(x) = \dfrac{2x - 1}{x + 1}$ is a bijection from $\mathbb{R} \setminus \{-1\}$ to $\mathbb{R} \setminus \{2\}$

   First observe that f is injective. If

   $$\frac{2a - 1}{a + 1} = \frac{2b - 1}{b + 1} \Rightarrow 2ab - b + 2a - 1 = 2ab - a + 2b - 1 \Rightarrow a = b.$$

   We now check to see that f is surjective.

   To find the range of the mapping f, solve $y = f(x)$ for x in terms of y.

   $$y \frac{2x - 1}{x + 1} = \Rightarrow y(x + 1) = 2x - 1 \Rightarrow 1 + y = 2x - xy \Rightarrow 1 + y = x(2 - y)$$

   $$\Rightarrow \frac{1 + y}{2 - y} = x$$

   which makes sense if $y \neq 2$.

   Therefore f is bijective.

Dfienition:

Let $f: S \to T$ be a mapping with domain S and range $R(f) \subseteq T$.

If E is a subset of S, then the *direct image* of E under f is the subset of T given by:

$$f(E) := \{f(x) : x \in E\}.$$

If H is a subset of T then the *inverse image* of H under f is the subset of S:

$$f^{-1}(H) := \{x \in S : f(x) \in H\}$$

□

When dealing with mappings it is important to remember which set is the domain and which is the codomain. Don't get confused about where the mapping is coming from and going to.

Example:

Let $f :\to \mathbb{Z}$ be defined by $f(x) = x^2$

- If $A = \{-3, -2, -1, 0, 1, 2, 3\}$ then $f(A) = \{0, 1, 4, 9\}$.
- If $A = \{-3, -2, -1, 0, 1, 2, 3\}$ then $f^{-1}(A) = \{-1, 0, 1\}$. Due to the sets that we are using, there is no way to take square roots or have complex numbers.
- If $B = \{-3, -2, -1\}$ then $f(B) = \{1, 4, 9\}$
- If $B = \{-3, -2, -1\}$ then $f^{-1}(B) = \emptyset$. Again, because we cannot take square roots of negative numbers with the sets that we are working with.

□

We need to be careful with our notation here. Despite the presence of a symbol that looks like an inverse, it is not saying that there is an inverse mapping.

Example:

Let $f :\to \mathbb{Z}$ be the mapping defined by $f(x) = x^2$ with $B = \{-3, -2, -1\}$.

- $f(f^{-1}(B)) = f(\emptyset) = \emptyset$. So $f(f^{-1}(B)) \neq B$.
- $f^{-1}(f(B)) = f^{-1}(\{1, 4, 9\}) = \{-3, -2, -1, 1, 2, 3\}$. So $f^{-1}(f(B)) \neq B$.

□

Definition: If $f : S \to T$ is a mapping and if $S_1 \subset S$, we can define a mapping $f_1 : S_1 \to T$ by

$$f_1(x) := f(x) \qquad \text{for} \quad x \in S_1.$$

The mapping f_1 is called the *restriction of* f *to* S_1. We have essentially thrown out part of the domain of the original mapping.

Example: Recall that above the mapping $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ was not one-to-one. Restrict the domain S_1 to the positive real numbers $\{x \in \mathbb{R} : x \geq 0\}$. The mapping $f_1 : S_1 \to \mathbb{R}$ is now a one-to-one mapping. If $f_1(x_1) = f_1(x_2)$, then $x_1^2 = x_2^2$ and since x_1 and x_2 are both nonnegative (remember, we have restricted the domain), this implies that $x_1 = x_2$.

□

1.4.2 Equivalence Relations

Many times in mathematics we want to talk about objects being the same. But this notion of "sameness" needs some clarification. It is clear that there are ways to say that $\sqrt{36}, 12/2,$ and $3!$ are all different ways to represent the same value of 6. What is not so clear is that there is a way to consider the numbers $-4, 6, 21,$ and 101 as being the same. What we need is a definition to talk about this idea precisely. When we say that $12/2 = 6$ we are really making a statement about the symbol $=$ and what relationship it establishes between the values on the right and left of the equals sign.

We will formalize this idea of sameness by looking at another particular type of relation. This will be a relation between a set S and itself with some extra conditions.

Definition: An *equivalence relation,* denoted $\triangleright$ on a set S is a relation from S to itself that satisfies these three properties for all $x, y, z \in S$:

1. Reflexive: $x \triangleright x$
2. Symmetric: If $x \triangleright y$, then $y \triangleright x$
3. Transitive: If $x \triangleright y$ and $y \triangleright z$ then $x \triangleright z$

□

Example: The clearest example, but one which our familiarity obscures the importance of the definition is equality $=$. We say that $x = y$ if the numerical value of x is the same as *y*.

It is clear that $x = x$. Furthermore if $x = y$ then $y = x$. Transitivity also follows since if $x = y$ and $y = z$ then $x = z$.

□

Example: Let T be the set of all triangles. For any two triangles $x, y \in T$ we say that *x* is similar to y if the three angle measures of x are equal to the angle measures of y.

x is similar to *x* as the angle measures of a triangle are equal to itself.

If *x* is similar to *y* then the three angle measures of *x* are equal to the angle measures of *y*. This means that the three angle measures of *y* are equal to the three angle measures of *x*, and so *y* is similar to *x*.

If *x* is similar to *y* and *y* is similar to z, then the three angle measures of x are equal to those of *y*, which are equal to those of z. Therefore the angle measures of x are equal to the angle measures of z and so x is similar to z.

Thus similarity is an equivalence relation. The sameness that we are identifying here only pertains to angle measures, not side lengths.

□

Example: From linear algebra, matrix A is row equivalent to matrix B if B is obtained from A by a finite number of elementary row operations. We will show that row equivalence is an equivalence relation.

Row equivalence is reflexive. A is row equivalent to itself as we can start with A, perform no row operations, and end with A.

Row equivalence is symmetric. If A is row equivalent to B then we obtain B from A by a finite number of elementary row operations. Each of these row operations can be reversed by an elementary row operation. Thus we can obtain A from B by a finite number of elementary row operations, and A is row equivalent to B.

Finally, row equivalence is transitive. If A is row equivalent to B and B is row equivalent to C then B can be obtained from A by a finite number of elementary row operations and C can be obtained from B by a finite number of elementary row operations. Thus C can be obtained from A by a finite number of elementary row operations and A is row equivalent to C.

□

Example: Suppose we say that for any real numbers x, y that $x \triangleright y$ if $|x - y| \leq 3$. Is this an equivalence relation?

It is true that $x \triangleright x$. Since $|x - x| = 0 < 3$, we know that this relation is reflexive. It is also true that this relation is symmetric. If $x \triangleright y$ then $|x - y| \leq 3$. It follows that $|y - x| \leq 3$ and so $y \triangleright x$.

However this relation is not an equivalence relation as it is not transitive. This can be seen by the following counterexample. Since $|2 - (-1)| \leq 3$ we see that $2 \triangleright -1$. Since $|-1 - (-3)| \leq 3$ it follows that $-1 \triangleright -3$. However it is not true that $2 \triangleright -3$ as $|2 - (-3)| = 5 > 3$.

□

The next example of an equivalence relation is one which we will come back to over and over again in our study of abstract algebra.

Example: Let $x, y \in \mathbb{Z}$. We say that $x \triangleright y$ if there exists an integer k such that $x = y + 5k$.

We see that $x \triangleright x$ since $x = x + 5 \cdot 0$.

If $x \triangleright y$ then $x = y + 5k$ for some integer k. By basic algebra we have $y = x + 5(-k)$. Since $-k$ is an integer this shows that $y \triangleright x$.

If $x \triangleright y$ and $y \triangleright z$ then there exist integers k, m such that $x = y + 5k$ and $z = y + 5m$. We again use some basic algebra to see that $x = (z - 5m) + 5k = z + 5(k - m)$. Since $k - m \in \mathbb{Z}$ we see that $x \triangleright z$.

□

This last example explains why we can consider $-4, 6$ and 101 to be the same. All of these numbers are equivalent by the above equivalence relation. Specifically, each of these numbers has a remainder of 1 when divided by 5. This equivalence relation is a special case of one so important that it is given a special name.

Definition: For integers *x,y,n*, we say *x is equivalent to y modulo n* if there exists an integer *k* such that $x = y + nk$. We denote this equivalence relation $x \triangleright y$ by $x = y \bmod n$.

□

Equivalence relations are defined on a particular set and partition this set into several subsets. These subsets are mutually disjoint. If we examine one of these subsets, every element contained therein is equivalent to every other element in the subset. This is the idea of an equivalence class.

Definition: Given a set *S* and element $x \in S$ with equivalence relation $\triangleright$, the *equivalence class of x* is the subset of *S* that contains all elements of *S* that are equivalent to *x*.

□

Example: Consider the equivalence relation on $\mathbb{Z}$ denoted by $x = y \mod 5$. This equivalence relation partitions the elements of $\mathbb{Z}$ into five equivalence classes:

- $\{\cdots, -10, -5, 0, 5, 10, \cdots\}$ Each of these elements x are of the form $x = 0 + 5 \cdot k$.
- $\{\cdots, -9, -4, 1, 6, 11 \cdots\}$ Each of these elements x are of the form $x = 1 + 5 \cdot k$.
- $\{\cdots, -8, -3, 2, 7, 12, \cdots\}$ Each of these elements x are of the form $x = 2 + 5 \cdot k$.
- $\{\cdots, -7, -2, 3, 8, 13, \cdots\}$ Each of these elements x are of the form $x = 3 + 5 \cdot k$.
- $\{\cdots, -6, -1, 4, 9, 14, \cdots\}$ Each of these elements x are of the form $x = 4 + 5 \cdot k$.

Every integer is in one and only one of these subsets. It is relatively easy to see from this example that there will be n equivalence classes from the modulo n equivalence relation, and these correspond to the remainders possible $(0, 1, \cdots, n-1)$ from division by n.

□

1.4.3 Exercises

1. Give an example of the following types of mappings $f : \mathbb{Z} \to \mathbb{Z}$:
   a) Injective but not surjective.
   b) Surjective but not injective.
   c) Neither surjective nor injective.
   d) Bijective.
2. Let $f : S \to T$ be a mapping and A, B subsets of *S*. Prove or give a counterexample:
   a) $f(A \cup B) = f(A) \cup f(B)$
   b) $f(A \cap B) = f(A) \cap f(B)$
   c) $B \subseteq f^{-1}(f(B))$
   d) $f(A) \backslash f(B) = f(A \backslash B)$
3. Let *X* be a finite set with n elements.
   a) How many elements are in $X \times X$?
   b) How many relations are there from *X* to *X*?
   c) How many mappings are there from *X* to *X*?
   d) How many equivalence relations are there from *X* to *X*?
   e) How many equivalence relations from *X* to *X* are also mappings?
4. Let $\subseteq$ denote subset inclusion, i.e. $A \subseteq B$ if *A* is a subset of *B*. Show that this relation is reflexive and transitive, but not symmetric and hence not an equivalence relation.
5. Prove that $\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x = y + nk, k \in \mathbb{Z}\}$ is an equivalence relation. (i.e. prove that $x \triangleright y$ by $x = y \mod n$ is an equivalence relation.)
6. On the set $\mathbb{R}$ define the relation $x \triangleright y$ if $|x| = |y|$. Is this relation an equivalence relation?
7. For the integers *x,y*, we say that $x = y \mod 6$ if there is an integer *k* such that$x = y+6k$. What are the equivalence classes for this equivalence relation?

2 Group Theory

2.1 Binary Operations

Of all of the abstract structures that we will study, the first of these is a group. Historically groups were among the first algebraic objects to be formally studied and are used in the definition of subsequent structures. Before presenting the definition of groups, the topic of binary operations must be explored. Suppose you saw the following things written on a wall:

red, circle . . . square

red, square . . . blue

square, red . . . blue

What is going on here? It's hard to tell exactly. We can see that for pair of elements from the set $\{\text{circle}, \text{blue}, \text{square}, \text{red}\}$ a third is mentioned. It's unclear what connection the third element has with the first two, but it appears that some rule dictates what happens. What does this situation have to do with abstract algebra? It's actually one that you have encountered before. A more familiar example of the above phenomenon is:

$8, 7 \cdots 15$

$2, 3 \cdots 5$

$12, 2 \cdots 14$

It's easy to see whats going on in this situation. Here we take two elements of the set $\mathbb{N}$ and return another element of the set $\mathbb{N}$. We can see that the third number is the sum of the first two. So we understand the rule that is operating here. But the overall structure of what is happening is identical to the first example. In each case we assign an element of a set to every pair of elements from that same set. If we think formally in terms of chapter 1, we can see that there is a mapping and Cartesian product at work here.

Definition: A **binary operation** μ on a set *S* is a mapping from the Cartesian product $S \times S$ into *S*. For each $(s,t) \in S \times S$ we will denote the element $\mu(s,t) \in S$ by $s \cdot t$

□

Note: In order for μ to be a binary operation on the set S the following must happen:

1. For every pair $(s,t) \in S \times S$ exactly one element is assigned.
2. The element assigned to the pair is also in *S*.

Notation: Typically the $s \cdot t$ notation is used when writing out binary operations. Although it is techically correct to write a binary operation as a mapping, this notation can get in the way of intuition. Typically the notion that a binary operation is a mapping is suppressed by using a symbol such as $\cdot$ and thinking of the binary operation as a type of multiplication. Despite the fact that $\cdot$ typically denotes standard multiplication of numbers, we can allow this to represent any binary operation.

□

Example:

The following are examples of binary operations on particular sets:

- Addition $+$ is a binary operation on each of the sets: $\mathbb{R}, \mathbb{Z}, \mathbb{N}$. We typically don't write $+(1,5) = 6$ to indicate $1 + 5 = 6$. Had we defined $\cdot(s,t) = s + t$ it would be appropriate to write $2 \cdot 5 = 7$, because our symbol $\cdot$ now represents addition.
- Addition $+$ is a binary operation on the set $\mathbb{C}$ of complex numbers. This addition is defined by $(a + bi) + (c + di) = (a + c) + (b + d)i$
- Standard multiplication $\cdot$ is a different binary operation on each of the sets $\mathbb{R}, \mathbb{Z}, \mathbb{N}$.
- $a \cdot b = ab - 2b$ is a binary operation on $\mathbb{Z}$.
- Let A, B be $n \times n$ matrices with real entries. $A \cdot B = (\det A - \det B)A$ is a binary operation.

□

We see from these examples that a set is not limited to one binary operation. One set can support a multitude of binary operations. The actual rule that determines a binary operation is really up to us, as long as it satisfies the definition. We must be on the lookout for situations such as the following.

Example: Let *D* denote the set of odd integers. Is ordinary addition a binary operation on *D*?

Even though for every pair of integers ordinary addition produces one integer, this is not a binary operation on the set *D*. The reason why is that the sum of any two odd numbers is even, which is not an element of the set *D*. What we have here is a mapping, but it is a mapping from $D \times D \to D^C$, not $D \times D \to D$.

□

Denition: A binary operation $\cdot$ is **associative** on S if for every $a, b, c \in S$ we have $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

□

Denition: A binary operation $\cdot$ is **commutative** on S if for every $a, b \in S$ we have $a \cdot b = b \cdot a$.

□

The definitions of commutative and associative binary operations are independent of one another. A binary operation can possess one, both, or neither of these properties.

- Addition on is $\mathbb{Z}$ both commutative and associative.
- Multiplication on is $\mathbb{Z}$ both commutative and associative.
- Subtraction is neither associative nor commutative on $\mathbb{Z}$. We see $(2-3)-4 = -5 \neq 3 = 2-(3-4)$, so it is not associative. The commutative property fails as well due to $2-3 \neq 3-2$.
- Let $M_n(\mathbb{R})$ denote $n \times n$ matrices with real entries. Matrix addition is commutative and associative, matrix multiplication is associative but not commutative.
- Given real numbers $x, y \in \mathbb{R}$, define a binary operation $x \cdot y = (x+y)^2$. Since

  $$(x+y)^2 = x^2 + 2xy + y^2 = y^2 + 2yx + x^2 = (y+x)^2 = y \cdot x$$

  this binary operation is commutative. However, this binary operation is not associative, as can be seen by comparing $(1 \cdot 2) \cdot 2 = 9 \cdot 2 = 121$ with $1 \cdot (2 \cdot 2) = 1 \cdot 16 = 289$

□

2.1.1 Exercises

1. Is the mapping defined by $a \cdot b = a/b$ a binary operation on the set $\mathbb{R}$? Explain.
2. Is the mapping defined by $a \cdot b = a \pm b$ a binary operation on the set $\mathbb{Q}$? Explain.

3. Is the mapping defined by $a \cdot b = a + b$ a binary operation on the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$? Explain.
4. Define the binary operation $x \cdot y = xy - 3$ on the set $\mathbb{Q}$. Is $\cdot$ associative? Is it commutative?
5. If $a \cdot b = b \cdot a$ for $a, b \in S$, is it true that $\cdot$ is commutative on the set S? Explain.
6. Let $\cdot$ be an associative and commutative binary operation on the set S. Let $A = \{s \in S \mid a \cdot a = a\}$. Prove that H is closed under $\cdot$.

2.2 Introduction to Groups

The group structure is important because it describes much of the mathematics that we have encountered as well as more advanced topics. Topics as diverse as addition of integers, multiplication of nonzero rational numbers, matrix multiplication of 3×3 matrices with real entries and nonzero determinant, and much more can all have the features of the mathematical object known as a group.

2.2.1 Basic Definitions

Definition: A group $\{G, \cdot\}$ is a nonempty set G closed under a binary operation $\cdot$ such that the following axioms are satisfied:

1. Associativity of $\cdot$: for all $a, b, c \in G$,

$$(a \cdot b) \cdot c = a \cdot (b \cdot c).$$

2. Identity element: There is an element $e \in G$ such that for all $g \in G$

$$e \cdot g = g \cdot e = g.$$

3. Inverse: For each $g \in G$ there exists an inverse $g^{-1} \in G$ such that

$$g \cdot g^{-1} = g^{-1} \cdot g = e.$$

□

There are a few things to mention about this definition. The first is that the set we are working with must be nonempty. So there has to be at least one element in our group. The next item of note is that while a group must have a binary operation that is associative, there is no mention about commutativity. This is why some of the formulas in the definitions appear redundant. If we do not have a commutative binary operation, we do need to be concerned about the order of the elements. That is why the definition of a group includes both $e \cdot g$ and $g \cdot e$. Of course, there are groups with commutative structures. To distinguish this feature we have an additional bit of terminology.

Definition: A group is **abelian** if its binary operation is commutative. A group is *nonabelian* if its binary operation is not commutative.

□

Definition: A *finite group* is a group with a finite number of elements. The *order* of a group G is the number of elements in the group, and is denoted by $|G|$.

□

Notation: There are a few remarks to make about the notation used in a group. The definition above employs a notation that suggests multiplication. Hence the inverse is written as we would typically think to write the multiplicative inverse of a real number. When the context is clear we will sometimes suppress the notation for the binary operation, writing xy rather than $x \cdot y$. We will also employ exponent notation, i.e., $x^2 = x \cdot x$, when this is convenient. There are times when it is more natural to use additive notation for our group operation. In this case the inverse of the element a is denoted by $-a$.

□

2.2.2 Examples of Groups

To see how far ranging the definition of a group is, we will look at an extensive series of examples of groups. Some of what follows constitute counterexamples. These are examples of sets with binary operation that fail to meet some part of the definition of a group.

Example:

1. $\mathbb{Z}$ under addition is an abelian group. Addition is associative, 0 is an identity, and for every $x \in \mathbb{Z}$ we have $-x$ as an inverse. Since $\mathbb{Z}$ satisfies these axioms, it forms a group under addition. Moreover, the addition is commutative so the group is abelian.

2. The set of positive integers $\mathbb{Z}^+$ under addition is not a group. Although the associative property holds, we do not have an identity. There is also no inverse element for any element in the group.

3. $\mathbb{Z}$ under multiplication is not a group. Although we have associativity of multiplication and 1 is the identity, any integers other than ± 1 do not have inverses. While it is true that $\frac{1}{2} \cdot 2 = 1$, the number $\frac{1}{2}$ is not an integer, and cannot be used as an inverse element of 2.

4. $n \times n$ matrices with real entries and nonzero determinant under matrix multiplication is a nonabelian group. Matrix multiplication is associative. The identity matrix I_n, a matrix with entry of 1 along the diagonal and zeros elsewhere, has the property that $A \cdot I_n = I_n \cdot A$ for all $n \times n$ matrices A. Since any matrix A in this set has nonzero determinant, there is an inverse matrix A^{-1}. Matrix multiplication is not commutative, so this group is nonabelian.

5. $\mathbb{Q}$ under multiplication is not a group. Although multiplication is associative on this set and 1 is the identity, the element 0 does not have an inverse. There is no rational number r such that $0 \cdot r = 1$.

6. $\mathbb{Q}^*$ – the set of nonzero rational numbers – under multiplication is an abelian group. The problem with the last example has been resolved and every element of $\mathbb{Q}^*$, which we may express as $\frac{p}{q}$, has inverse $\frac{q}{p}$.

7. The set $\{e\}$, with binary operation $e \cdot e = e$ is an abelian group. Associativity follows by checking $(e \cdot e) \cdot e = e \cdot e = e \cdot (e \cdot e)$. Since $e \cdot e = e$ this one element is its own identity and inverse.

8. The set $\{e, g\}$, with binary operation:

   - $e \cdot e = e$
   - $e \cdot g = g$
   - $g \cdot e = g$
   - $g \cdot g = e$

is an abelian group. Associativity is a bit tedious to verify. But if we check, we will find that associativity holds for all eight cases that are possible. The element e is an identity. Each element is its own inverse. Since the binary operation is commutative, this is an abelian group.

9. The complex n th roots of unity is is the set of all complex numbers z such that $z^n = 1 + 0i$. This set forms an abelian group under multiplication of complex numbers. Multiplication of complex numbers is associative and commutative. The number $1 + 0i$ is a complex n th root of unity as $(1 + 0i)^n = 1 + 0i$ and this serves as the identity since $(1 + 0i) \cdot (a + bi) = (a + bi)$ for any complex number. Furthermore, every n th root of unity has a multiplicative inverse. Suppose $z = a + bi$ and $z^n = 1$. Let $\bar{z} = \frac{a-bi}{a^2+b^2}$ and we have
$$z \cdot \bar{z} = (a + bi)\frac{a - bi}{a^2 + b^2} = \frac{(a + bi)(a - bi)}{a^2 + b^2} = \frac{a^2 + abi - abi - i^2b}{a^2 + b^2} = \frac{a^2 + b^2}{a^2 + b^2} = 1 + 0i.$$
While this shows that $\bar{z}$ is the inverse of z, we still need to show that $\bar{z}$ is itself an n th root of unity:
$$\bar{z}^n = (1 + 0i) \cdot \bar{z}^n = z^n \cdot \bar{z}^n = (z \cdot \bar{z})^n = (1 + 0i)^n = 1 + 0i.$$

10. The set $\mathbb{Z}_n$ of equivalence classes modulo n , which we denote $\{[0], [1], [2], \cdots , [n - 1]\}$ forms a group under the addition $[x] + [y] = [(x + y) \bmod n]$. Associativity of addition is inherited from associativity of addition in $\mathbb{Z}$. The equivalence class associated to $[0]$ is the identity. For any equivalence class $[x]$, consider $[n - [x] \bmod n]$. Since $[x] + [n - [x] \bmod n] = [n \bmod n] = [0]$, every element in $_n$ has an inverse.

11. To see how the last example works for a specific value of n , we will look at $\mathbb{Z}_3$.

    We will denote equivalence classes

    - $[0] = \{\cdots , -6, -3, 0, 3, 6, \cdots\}$
    - $[1] = \{\cdots , -5, -2, 1, 4, 7, \cdots\}$
    - $[2] = \{\cdots , -4, -1, 2, 5, 8, \cdots\}$

    The binary operation is defined as:

    - $[0] + [0] = [0]$
    - $[1] + [0] = [0] + [1] = [1]$
    - $[2] + [0] = [0] + [2] = [2]$
    - $[1] + [1] = [2]$
    - $[1] + [2] = [2] + [1] = [0]$
    - $[2] + [2] = [1]$

It has been noted that associativity has been inherited from $\mathbb{Z}$. Since $a + (b + c) = (a + b) + c$ in $\mathbb{Z}$, when we work modulo n, this is also true. $[0]$ is the identity element. For inverses $[0]$ is its own inverse and $[1]$ and $[2]$ are inverses of each other. Although the type of addition seen in this example may seem unnatural, we use it every day when we look at a clock. Just as $5 + 10 = 3 \mod 12$, five hours after 10 o'clock is 3 o'clock. This type of example will be a major one that we will continually return to throughout the book.

□

We see from the above examples that the group structure is very much dependent on both the set we are using as well as the binary operation on the set. It should be clear from the variety displayed here, that we are really just scratching the surface of the number of situations that can be described in terms of a group.

2.2.3 Basic Theorems Regarding Groups

Now that we've studied several examples of groups, we'll look at what we can prove concerning this definition.

Theorem 4. *The identity element G of a group is unique.*

Proof. As before for a uniqueness proof, we will begin by assuming that the identity is not unique. Suppose by way of contradiction that e, f are both identity elements of G and that $e \neq f$.

However, $e = f \cdot e$ since f is an identity and $f \cdot e = f$ since e is an identity. Combining these equalities we see $e = f \cdot e = f$. Our supposition is false and the identity element of a group is unique.

□

Note:

The hypothesis that we are working with a group is actually a little more than we need. Nothing in the above proof required any part of the definition of a group. All that was assumed is that we had an identity element. The above proof could be used for any set S and binary operation $\cdot$ for which there is an identity element e.

Identities are not the only things in groups that are unique.

Theorem 5. *For the group G, the inverse of an element $g \in G$ is unique.*

Proof. Suppose by way of contradiction that $g' \neq g''$ are both inverses of g.

$$g'' = e \cdot g'' = (g' \cdot g) \cdot g'' = g' \cdot (g \cdot g'') = g' \cdot e = g'$$

This shows that $g'' = g'$, therefore inverses are unique.

□

The previous theorem shows that for a group G and $a, b, x \in G$, the equation $ax = b$ has a unique solution for x. Since a has a unique inverse, a^{-1} we may write $x = a^{-1}ax = a^{-1}b$.

Theorem 6. *If a, b are elements of a group G then $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$.*

Proof. We prove this by using the definition of inverse. For a group G and $a, b \in G$, we consider

$$(a \cdot b) \cdot (b^{-1}a^{-1}) = a \cdot ((b \cdot b^{-1}) \cdot a^{-1}) = a \cdot (e \cdot a^{-1}) = a \cdot a^{-1} = e$$

A similar series of steps shows that $(b^{-1} \cdot a^{-1}) \cdot (a \cdot b) = e$. Therefore $b^{-1} \cdot a^{-1} = (a \cdot b)^{-1}$.

□

WARNING: We need to be careful about how we prove statements regarding groups. Much of the algebra that we are accustomed to may not be valid, because it relies on the assumption of the commutative property. For instance, it's not always true anymore that $(xy)^2 = x^2y^2$. This is because in a nonabelian group $(x \cdot y) \cdot (x \cdot y) = x \cdot (y \cdot x) \cdot y$, but there is no justification that we have to switch the order of the $y \cdot x$ to $x \cdot y$. We need to exercise caution in how we use cancellation on two sides of an equation. If we "multiply" by an element on the left side of an equation, we must multiply the same element on the right side of the equation.

□

For illustration of how to prove statements regarding groups, here are a couple of basic proofs.

Theorem 7. *G is an abelian group if and only if* $(xy)^2 = x^2y^2$ *for all elements* $x, y \in G$.

This is an "if and only if proof," so there are really two statements to prove.

Proof. Suppose that *G* is an abelian group. We consider

$$(x \cdot y)^2 = (x \cdot y) \cdot (x \cdot y) = x \cdot (y \cdot x) \cdot y = x \cdot (x \cdot y) \cdot y = (x \cdot x) \cdot (y \cdot y) = x^2 \cdot y^2$$

We are justified in making the statement $x \cdot y = y \cdot x$ due to the hypothesis that *G* is abelian. Thus we have shown "If *G* is abelian, then $(x \cdot y)^2 = x^2 \cdot y^2$." It remains to show the other half of the statement.

Now suppose that for all $x, y \in G$ we have $(xy)^2 = x^2y^2$. Expanding this expression we see $x \cdot y \cdot x \cdot y = x \cdot x \cdot y \cdot y$. If we multiply on the left by x^{-1} and on the right by y^{-1} we have:

$$x^{-1} \cdot x \cdot y \cdot x \cdot y \cdot y^{-1} = x^{-1} \cdot x \cdot x \cdot y \cdot y \cdot y^{-1}$$

which simplifies to $y \cdot x = x \cdot y$. Since this statement is true for all $x, y \in G$ we have shown that *G* is an abelian group. This shows "If $(xy)^2 = x^2y^2$ for all $x, y \in G$ then *G* is abelian."

Combined with the other part of the proof we have shown *G* is abelian if and only if $(xy)^2 = x^2y^2$ for all elements $x, y \in G$.

□

2.2.4 Group Tables

The last basic consideration regarding groups is their presentation. Rather than the cumbersome lists that we have used to show all possible binary operations, we can organize these lists into a table. These tables have the advantage of being compact, easy to read, and connect to the familiar notion of a multiplication table. To read the table, the first element of our binary operation comes from the leftmost column of the table. The second element of the binary operation comes from the top row. Where the row and column intersect is the product of the binary operation of these two elements. Group tables are better suited for work with finite groups of low order. Below we will see a few of these.

Example:

Above we saw an abelian group with two elements. This corresponded to the set $\{e, g\}$, with binary operation:

- $e \cdot e = e$
- $e \cdot g = g$
- $g \cdot e = g$
- $g \cdot g = e$

$\cdot$	e	g
e	e	g
g	g	e

To show that $g \cdot g = e$ we note that the g column and the g row intersect at the element e.

□

Example:

A larger group table, for a group with four elements is:

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

.

The 16 entries of the table define the binary operation.

There are a few features of note in a group table of a finite group of order n.

- Group tables will contain each of the n elements of *G* in each row and each column. This forms a sort of Sudoku puzzle in which no element can occur twice in the same row or same column.
- The same group can be expressed in terms of different tables, however the set of binary operations in the two tables will be the same.
- Abelian groups can be arranged in a way that is symmetric along the main diagonal running from the upper left to the lower right of the table.

2.2.5 Exercises

1. Construct a group table for a group with three elements $\{e, a, b\}$.
2. Construct a group table for a group with four elements $\{e, a, b, c\}$. Leaving the elements in this order on both the top row and left column, is there only one way to form a group table with four elements?
3. Prove that *G* is an abelian group if and only if $(a \cdot b)^{-1} = a^{-1}b^{-1}$
4. Given any set *A* the power set of *A*, denoted $P(A)$, is the set of all subsets of *A*. Prove that $P(A)$ with binary operation the symmetric difference is a group.
5. Prove that for any element a in the group G, $(a^{-1})^{-1} = a$.
6. Prove that for the group *G* and every $a_1, a_2, \cdots, a_n \in G$, $(a_1 \cdot a_2 \cdot \cdots \cdot a_n)^{-1} = a_n^{-1} \cdot \cdots a_2^{-1} \cdot a_1^{-1})$

7. Prove that for a group *G* and elements $c, d, y \in G$, the equation $yc = d$ has a unique solution for y.
8. Let G be a group where for every $x \in G$, $x \cdot x = e$. Prove that *G* is abelian.
9. Let $X = \mathbb{R} \cup \{-\infty\}$. We define a binary operation on X denoted $+$ as follows: if x and y are both elements in $\mathbb{R}$, then $x + y$ is the usual sum of $x + y$ (an element of R). If at least one of x or y is $-\infty$, then $x + y = -\infty$. Show that X with binary operation $+$ is commutative and associative, but not a group.
10. Prove that for the finite group *G* with identity e and order $2k$ for $k \in \mathbb{N}$, there is an element $g \neq e$ such that $g \cdot g = e$
11. An element of the group *G* is idempotent if $g \cdot g = g$. Prove that every group has one and only one idempotent element.

2.3 Cyclic Groups

We have already seen that the complex n th roots of unity C_n form a group under complex multiplication. If we plot all of these points in the complex plane, we see that they are all located on a unit circle. Adjacent points are separated by an angle of $2\pi/n$. A primitive n th root of unity is complex number $\xi = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$. For this complex number n is the smallest positive integer such that $\xi^n = 1 + 0i$. What this means is that $\xi, \xi^2, \xi^3, \cdots, \xi^{n-1}, \xi^n$ are all distinct n th roots of unity. Successive powers of ξ produce the entire group C_n .

For another instance of this same phenomena, consider the equivalence classes of the integers modulo n , written as $\mathbb{Z}_n$, under addition modulo n . The equivalence class 1 can be added to itself over and over: $1, 1 + 1 = 2, 1 + 1 + 1 = 3, \cdots, \overbrace{1 + 1 + \cdots 1}^{n} = 0$. In this way we obtain every element of $\mathbb{Z}_n$.

These groups can be classified by the fact that we can use one element repeatedly with the binary operation to produce the entire group. We need some notation to deal with this idea.

Notation:

For a group *G* and $a \in G$ we define the notation $\langle a \rangle := \{a^n | n \in\}$

□

Definition: An element a of the group *G* is a *generator* of *G* if $G = \langle a \rangle$. *G* is *cyclic* if there is some element a in *G* that generates *G*.

□

Example: $\mathbb{Z}_4$ is cyclic. We note that $3 + 3 = 2$, $3 + 3 + 3 = 1$, and $3 + 3 + 3 + 3 = 0$. So 3 generates $\mathbb{Z}_4$. The element 1 also generates $\mathbb{Z}_4$.

□

Denition:

Let a be an element of the group G. The *order of* a is the number of elements in the set $\langle a \rangle$.

□

Example:

The group $\mathbb{Z}_8$ is also cyclic.

- $\langle 0 \rangle = \{0\}$
- $\langle 1 \rangle = \{1, 2, 3, 4, 5, 6, 7, 0\}$
- $\langle 2 \rangle = \{2, 4, 6, 0\}$ This is obtained by adding 2 to itself repeatedly. Since our addition is modulo 8, $6 + 2 = 0$. The element 2 has order 4.
- $\langle 3 \rangle = \{3, 6, 1, 4, 7, 2, 5, 0\} = \langle 1 \rangle$.
- $\langle 4 \rangle = \{4, 0\}$. The element 4 has order 2.
- $\langle 5 \rangle = \{5, 2, 7, 4, 1, 6, 3, 0\} = \langle 1 \rangle$.
- $\langle 6 \rangle = \{6, 4, 2, 0\} = \langle 2 \rangle$
- $\langle 7 \rangle = \{7, 6, 5, 4, 3, 2, 1, 0\} = \langle 1 \rangle$.

We see that the elements 1,3,5, and 7 are all generators of the group $\mathbb{Z}_8$ as they all have order 8.

□

Example: The integers $\mathbb{Z}$ under standard addition forms a cyclic group. The generators are 1 and –1.

□

Theorem 8. *If G is a cyclic group, then it is abelian.*

Proof. Suppose G is cyclic. We know that it has a generator a. Thus every element $g \in G$ can be written as $a^k = g$ for some $k \in \mathbb{N}$. Let $x, y \in G$. There are integers m, n such that $x = a^m$ and $y = a^n$.

$$x \cdot y = a^m \cdot a^n = a^{m+n} = a^{n+m} = a^n \cdot a^m = y \cdot x.$$

Since $x \cdot y = y \cdot x$, the cyclic group G is abelian. □

Example: The converse of the last theorem is not true. Consider the group given by the group table:

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

This group is abelian, however it is not cyclic as there is no element that generates the entire group. This can be seen by noting that $a^2 = b^2 = c^2 = e$. This group is known as the Klein 4-group or Vierergruppe in German. This is why this group is commonly denoted by *V*.

□

Theorem 9. *If a is a generator of a finite cyclic group G of order n then the other generators of G are the elements of the form a^r where r is relatively prime to n.*

Proof. Let G be a cyclic group with generator a and order n. Suppose that r is relatively prime to n. We see that $\langle a^r \rangle = e, a^r, a^{2r}, \cdots, a^{(m-1)r}$ where $a^{mr} = e$ and m is the smallest such positive integer. Thus the order of the group n is divides mr. Since n and r are relatively prime, n divides m. Thus there are at least n distinct elements in the list $e, a^r, a^{2r}, \cdots, a^{(m-1)r}$. But there can only be n at most as each $a^{jr} \in G$. Therefore $\langle a^r \rangle = G$.

□

Example: Consider $G = \mathbb{Z}_{12}$ G is generated by $1, 5, 7, 11$. $\langle 2 \rangle = \{0, 2, 4, 6, 8, 10\}$

$\langle 3 \rangle = \{0, 3, 6, 9\} = \langle 9 \rangle$
$\langle 4 \rangle = \{0, 4, 8\} = \langle 8 \rangle$
$\langle 6 \rangle = \{0, 6\}$

□

Example: Consider $G = \mathbb{Z}_{24}$. G is generated by $1, 5, 7, 11, 13, 17, 19, 23$.

□

2.3.1 Exercises

1. What are the generators of the group $\mathbb{Z}_{60}$?
2. What is the order of the element 32 in the group $\mathbb{Z}_{56}$?
3. Find the number of generators of the group $\mathbb{Z}_{pq}$ where p and q are prime numbers.

2.4 Dihedral Groups

One powerful feature of the language of group theory is that it can be used to express symmetries. This makes our abstract study of groups very important for many subjects in the physical world. We will look at symmetries of regular polygons and see where this leads us. To be precise, a symmetry of a polygon is a self-congruence of the polygon. Even more precisely a congruence is a distance preserving one-to-one map of the polygon with itself. We can compose two symmetries and the result is another symmetry. The reason for this is that the composition of two one-to-one mapping is again a one-to-one mapping. Thus we can consider this composition of mappings as a binary operation that combines two symmetries into one.

Theorem 10. *The set of symmetries of a regular polygon, with binary operation defined as the composition of symmetries is a group.*

Proof. All that we need to do is check that this set and binary operation meet the definition of a group. The composition of any mappings is associative, and so the composition of symmetries is also associative. The identity mapping of a polygon serves as the identity symmetry. Finally, if a mapping is one-to-one, then an inverse exists. Thus any symmetry has an inverse.

□

Denition:

The *Dihedral group* D_n of order $2n$ is the group of symmetries of a regular n -gon.

□

Now that we know this is a group structure, we can investigate this more fully. Again the formalism of defining a symmetry as above gets in the way of our intuition. Recall that a regular polygon has n sides of equal length and n angles of equal measure. We will suppose this polygon is fixed in space. There are two types of things that we can do to the polygon and still preserve congruence. Rotational symmetries rotate the polygon counterclockwise about its center by $2\pi/n$ radians. Reflection symmetries operate by flipping the polygon across an axis or line of symmetry. Counting the identity as a rotation of 0 radians, there are n rotational symmetries. There are also n reflection symmetries. This means that the order of the group D_n is $2n$.

Note: The notation for a dihedral group varies from textbook to textbook. Since the group has $2n$ elements – n rotational symmetries and n reflection symmetries ï¿½ some books use D_{2n} to indicate our D_n. Technically D_1 and D_2 can be defined using an alternate definition. However, this notation is rarely used and we will see that it will be unnecessary to consider these two groups.

□

Example:

We will look at D_3, the symmetries of an equilateral triangle. We can rotate by 0 radians, rotate by $2\pi/3$ radians, or rotate by $4\pi/3$ radians and the result is a triangle with the same orientation. If we rotate by $6\pi/3 = 2\pi$ radians, this is equivalent to no rotation at all. We note that rotation by $2\pi/3$ twice is the same as rotation by $4\pi/3$. To summarize all of this information, we let r_1 denote the rotation by $2\pi/3$, r_2 rotation by $4\pi/3$ and e the identity of no rotation.

This gives us:

$\cdot$	e	r	r^2
e	e	r_1	r_2
r_1	r_1	r_2	e
r_2	r_2	e	r_1

Now we look at the reflection symmetries. We will denote these by u, d, v , corresponding to "up," "down," and "vertical" for the three possible directions of axes of symmetry. We see very quickly that these reflection symmetries behave differently than the rotational symmetries. For starters, the composition of a reflection with itself is the identity $u^2 = v^2 = d^2 = e$. If we compose two different reflections we see that this is equivalent to a rotation. For instance $u \cdot v = r_1$ and $v \cdot u = r_2$.

$\cdot$	e	r_1	r_2	v	u	d
e	e	r_1	r_2	v	u	d
r_1	r_1	r_2	e	u	d	v
r_2	r_2	e	r_1	d	v	u
v	v	d	u	e	r_2	r_1
u	u	v	d	r_1	e	r_2
d	d	u	v	r_2	r_1	e

To summarize, D_3 is a group of order 6 and is a nonabelian group.

□

Notice that the dihedral group can be nonabelian. Although there are some elements that commute, such as r_1 and r_2, there are pairs such as u, v that do not. Dihedral groups are important because they are one of the more straightforward examples of nonabelian groups. They provide an example that is easy to think about and even manipulate by hand.

2.4.1 Alternate Definition

We may also define a dihedral group in terms of two generators. In the above example of D_3 we could have denoted $r_1 = r$ and $r_2 = r^2$. Furthermore, we could have chosen any of the reflections, such as v and observed that v, vr, vr^2 are all different reflection symmetries. Keeping in mind that $v^2 = r^3 = e$ we have the following table:

$\cdot$	e	r	r^2	v	vr	vr^2
e	e	r	r^2	v	vr	vr^2
r	r	r^2	e			
r^2	r^2	e	r			
v	v	vr	vr^2	e	r	r^2
vr	vr	vr^2	v			
vr^2	vr^2	v	vr			

There are some gaps in the table, caused by problems such as what element rv is equal to. By examining the symmetries we see that $rv = vr^2$. This additional piece of information is enough to fill in the rest of the table:

$\cdot$	e	r	r^2	v	vr	vr^2
e	e	r	r^2	v	vr	vr^2
r	r	r^2	e	vr^2	v	vr
r^2	r^2	e	r	vr	vr^2	v
v	v	vr	vr^2	e	r	r^2
vr	vr	vr^2	v	r^2	e	r
vr^2	vr^2	v	vr	r	r^2	e

This points the way to an alternate definition of the dihedral group:

Definition: The *dihedral group* $D_n = \{v^i r^j \mid i, j \in \mathbb{Z},\ r^n = v^2 = e, rv = vr^{n-1}\}$

□

2.4.2 Exercises

1. For a square let's consider D_4. There are four rotational symmetries: e, r_1, r_2, r_3 where r_n denoting rotation clockwise by $\frac{\pi}{2}n$ radians, h a reflection across a horizontal line v a reflection across a vertical line, u an upward diagonal axis of reflection, d a downward diagonal axis of reflection. Complete the following group table:

$*$	e	r_1	r_2	r_3	h	v	u	d
e	e	r_1	r_2	r_3	h	v	u	d
r_1	r_1	r_2	r_3	e				
r_2	r_2	r_3	e	r_1				
r_3	r_3	e	r_1	r_2				
h								
v								
u								
d								

2. Prove that D_n is nonabelian for $n \geq 3$.
3. Prove that D_n is noncyclic for $n \geq 3$.

2.5 Groups of Permutations

In addition to cyclic and dihedral groups, there are a number of other ways that a group can arise. The next that we will consider involves the concept of permutations of a finite set.

Definition: A *permutation* of a set *A* is a mapping $\phi : A \to A$ that is both one to one and onto.

□

Since we are going to define a group based upon these permutations, we need to have a binary operation to use with them. Similar to composing two symmetries together to obtain another symmetry, we can compose two permutations together and obtain another permutation. The main importance of the following theorem is that it establishes that $\circ$ is a binary operation on the set of permutations of the set *A*.

Theorem 11. *Given two permutations, σ, τ the operation $\sigma \circ \tau$ formed by composition of mappings is also a permutation.*

Proof. Suppose $\sigma\tau(a_1) = \sigma\tau(a_2)$. Since σ is one-to-one, $\tau(a_1) = \tau(a_2)$. Since τ is one-to-one, $a_1 = a_2$. Therefore $\sigma\tau$ is one-to-one.

Choose $a \in A$. Since σ is onto, there exists $a' \in A$ with $a = \sigma(a')$. Since τ is onto, there exists $a'' \in A$ with $a' = \tau(a'')$. Therefore $a = \sigma(a') = \sigma(\tau(a''))$ and $\sigma\tau$ is onto.

We have shown that if σ, τ are permuations, then $\sigma \circ \tau$ is also a permutation.

□

Example:

Suppose $A = \{1, 2, 3, 4\}$. We will denote the permutation that maps as follows:

$$\sigma(1) = 3, \sigma(2) = 1, \sigma(3) = 4, \sigma(4) = 2$$

by the notation: $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$. The key to reading this notation is that the image of each of the elements in the top row is located directly below. The 3 is below 1 because $\sigma(1) = 3$. Since there are 4 elements in the set A there are $4! = 24$ permutations. One permutation that is different than σ is given by

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$$

We may now compose these permutations just as we compose two mappings. For instance $\sigma \circ \tau(2) = \sigma(\tau(2)) = \sigma(1) = 3$. This means that the permutation of $\sigma\tau$ will have 3 directly below 2. The entire permutation is

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}$$

□

Theorem 12. *Let A be nonempty and S_A the collection of all permutations of A. Then S_A is a group under permutation multiplication.*

Proof. All that we need to do is check to see that the definition of a group is satisfied:

- The composition of mapping is associative, thus our binary operation is associative.
- The permutation $\iota(a) = a$ for all $a \in A$ is the identity, since for any permuation σ:
$$\sigma \circ \iota = \iota \circ \sigma$$
- Since a permutation is a one-to-one and onto mapping, it has an inverse mapping that is also one-to-one and onto. This is simply a permutation that reverses the order of the mapping σ. Thus we have an inverse σ^{-1}.

□

Definition: The permutations of a finite set with n elements is the *symmetric group* S_n and has $n!$ elements.

□

It is important to note how quickly the size of S_n increases as n increases. For instance the order of S_5 is 120, and the order of S_{10} is $3,628,800$. The factorial goes a long way. Due to the order of S_n for relatively small values of n, we will look at some symmetric groups of the lowest orders.

2.5.1 Examples of S_n

Example: For $n = 1$ there is one permutation from the set $A = \{1\}$ to A. This is the identity permutation, and so S_1 consists of $\{\iota\}$. It is clear that $\iota \circ \iota = \iota$.

□

Example: For $n = 2$ there are two permutations from the set $A = \{1, 2\}$ to A. One permutation is the identity permutation i and the other permutation, which we will call σ maps as follows: $\sigma(1) = 2$ and $\sigma(2) = 1$. In other notation $\sigma = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$. We may form a group multiplication table with these two permutations i and σ.

$\circ$	ι	σ
ι	ι	σ
σ	σ	ι

The only composition that needs some explaining is $\sigma \circ \sigma = \iota$. It helps to think that since σ switches 1 and 2, applying σ a second time will switch 1 and 2 back to their original positions.

□

Example: For $n = 3$ we now have $3! = 6$ permutations. We list the permutations of S_3 below:

$$\iota = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

$$\rho_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \rho_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \rho_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

There are $6^2 = 36$ possible pairings of these six permutations. Fortunately we do not have to calculate all of these. The $\iota \circ \mu = \mu \circ \iota = \mu$ for any permutaton μ. Furthermore, since each of the permutations labeled with ρ switch two elements of the set A, $\rho_1^2 = \rho_2^2 = \rho_3^2 = i$. The other compositions will take some work to figure out.

For $\rho_1 \circ \sigma_1$:

- $(\rho_1 \circ \sigma_1)(1) = \rho_1(\sigma_1(1)) = \rho_1(2) = 3$
- $(\rho_1 \circ \sigma_1)(2) = \rho_1(\sigma_1(2)) = \rho_1(3) = 2$
- $(\rho_1 \circ \sigma_1)(3) = \rho_1(\sigma_1(3)) = \rho_1(1) = 1$

We can see that $\rho_1 \circ \sigma_1 = \rho_2$

$\cdot$	ι	σ_1	σ_2	ρ_1	ρ_2	ρ_3
ι	ι	σ_1	σ_2	ρ_1	ρ_2	ρ_3
σ_1	σ_1	σ_2	ι	ρ_3	ρ_1	ρ_2
σ_2	σ_2	ι	σ_1	ρ_2	ρ_3	ρ_1
ρ_1	ρ_1	ρ_2	ρ_3	ι	σ_2	σ_1
ρ_2	ρ_2	ρ_3	ρ_1	σ_1	ι	σ_2
ρ_3	ρ_3	ρ_1	ρ_2	σ_2	σ_1	ι

There are a few features of note about the symmetric group S_3. We see that this group is not cyclic as there is no generator. This is nonabelian as well. This is clear by seeing that $\rho_1 \circ \sigma_1 \neq \sigma_1 \circ \rho_1$.

□

The symmetric group S_3 gives us an indication of the nature of S_n for most values of n. S_n is neither cyclic nor abelian for $n \geq 3$.

Even though the entire symmetric group is important, rather than looking at all permutations, we will only consider some of them. This partial set of permutations can be used to form a group.

Example: Give the multiplication table for the group generated by

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{pmatrix}.$$

By composing ρ with itself repeatedly, we see that $6 = n$ is the first positive integer such that $\rho^n = \iota$.

$\circ$	ι	ρ	ρ^2	ρ^3	ρ^4	ρ^5
ι	ι	ρ	ρ^2	ρ^3	ρ^4	ρ^5
ρ	ρ	ρ^2	ρ^3	ρ^4	ρ^5	ι
ρ^2	ρ^2	ρ^3	ρ^4	ρ^5	ι	ρ
ρ^3	ρ^3	ρ^4	ρ^5	ι	ρ	ρ^2
ρ^4	ρ^4	ρ^5	ι	ρ	ρ^2	ρ^3
ρ^5	ρ^5	ι	ρ	ρ^2	ρ^3	ρ^4

□

2.5.2 Exercises

1. Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 1 & 4 & 6 & 3 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 1 & 2 & 6 & 4 & 5 \end{pmatrix}$ in S_6.
   a) Calculate $\sigma \circ \tau$
   b) Calculate $\tau \circ \sigma$
   c) Calculate σ^{200}
   d) What is the order of τ?

2. Is there an element of order 6 in S_4? Explain.
3. Is there an element of order 6 in S_5? Explain.

2.6 Alternating Groups

The symmetric group on a set of n elements has an underlying set containing all $n!$ permutations of the set. We saw that we can form groups with only some of the permutations. We will now see that there is a particular group that can be formed by choosing exactly half of the permutations of S_n. We will need to introduce some new definitions to make it clear which half of the permutations we will be using.

Definition: Let a, b be elements of a finite set A and $\sigma \in S_n$. We say that $a \prec b$ if and only if there exists an $n \in \mathbb{Z}$ such that $\sigma^n(a) = b$ for some $n \in \mathbb{Z}$.

□

Theorem 13. *The relation defined above is an equivalence relation.*

Proof. We begin with $\sigma \in S_n$ and the relation $\prec$ as defined above. We must check that $\prec$ possesses the three conditions of an equivalence relation. Let $a, b, c \in A$

- The relation $\prec$ is reflexive. We see that $a \prec a$ since $a = \sigma^0(a) = \iota(a)$
- The relation $\prec$ is symmetric. If $a \prec b$ then $b = \sigma^n(a) \Rightarrow a = \sigma^{-n}(b)$ so $b \prec a$.
- The relation $\prec$ is transitive. Suppose $a \prec b$ and $b \prec c$. Then $b = \sigma^m(a)$ and $c = \sigma^n(b)$. Therefore $c = \sigma^{m+n}(a)$ and $a \prec c$.

□

Definition: The *orbits* of $\sigma \in S_n$ are the equivalence classes under the equivalence relation $\prec$ defined above.

□

Example:

What are the orbits of the permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 3 & 1 & 6 & 7 & 2 & 5 \end{pmatrix}$?

We start with any element and apply σ repeatedly. Since $\sigma(1) = 4$ this tells us that 1 and 4 are in the same orbit. We continue this process and see $\sigma(4) = 6, \sigma(6) = 2, \sigma(2) = 3, \sigma(3) = 1$. Since we have arrived back where we started, we are done. The set $\{1, 4, 6, 2, 3\}$ is an orbit. Now choose an element, such as 5, that is not in this orbit. We see that $\sigma(5) = 7$ and $\sigma(7) = 5$, so $\{5, 7\}$ is another orbit. There are no other elements to check, so we are done.

□

Definition:

We call a permutation $\sigma \in S_n$ a *cycle* if it there is at most one orbit with more than one element of *A*, meaning every element not in this orbit is fixed. The number of elements in this orbit is called the *length* of the cycle.

□

Example:

- The permutation from the previous example $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 3 & 1 & 6 & 7 & 2 & 5 \end{pmatrix}$ is not a cycle. There are two orbits with more than one element.
- The permutation $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 3 & 5 & 4 & 7 & 6 & 2 \end{pmatrix}$ is a cycle. The permutation has the orbit $\{2, 3, 5, 7\}$ and all other elements are fixed.
- The identity permutation ι is a cycle. Every element is fixed by this permutation, so every orbit has one element.

□

The notation for permutations that we have been using is called the tableau notation. Although this notation has its merits, it complicates some matter. We will introduce a different notation for a permutation. The advantage of this notation is its compactness and how it connects to our notion of cycles.

Notation:

Let $a_1, a_2, \cdots, a_n$ be the elements of the set A that are permuted by a cycle and let $b_1, b_2, \cdots, b_k$ be elements that are fixed. We express the permutation

$$\sigma = \begin{pmatrix} a_1 & a_2 & \cdots & a_n & b_1 & b_2 & \cdots & b_k \\ a_2 & a_3 & \cdots & a_1 & b_1 & b_2 & \cdots & b_k \end{pmatrix} = (a_1, a_2, a_3, \cdots, a_n)$$

This indicates that $\sigma(a_1) = a_2, \sigma(a_2) = a_3$ and so on. Any elements of A that do not appear in this cycle notation are assumed to be fixed by the permutation.

□

Theorem 14. *Every permutation σ of a finite set is formed from disjoint cycles.*

Proof. Let σ be a permutation of a finite set A. Partition A into the orbits of σ and denote these $A_1, A_2, \cdots, A_k$. Now let τ_i be the cycle that fixes every $a \notin A_i$ and $\tau_i(a) = \sigma(a)$ if $a \in A_i$. The A_i are disjoint, the cycles τ_i are also disjoint. Since $\tau_1\tau_2 \cdots \tau_k(a) = \sigma(a)$ for every $a \in A$, $\sigma = \tau_1\tau_2 \cdots \tau_k$.

Example: Write the permutation as a product of cycles: $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 5 & 1 & 6 & 8 & 7 \end{pmatrix}$.

First we determine the orbits of the permutation. They are $\{1,4,5\}, \{2,3\}, 6, \{7,8\}$. This permutation is composed of the cycles $(1,4,5), (2,3),$ and $(7,8)$. The element 6 is fixed. So we may write the permutation as $(1,4,5)(2,3)(7,8)$. These cycles are disjoint, and so they can be rearranged to give the same permutation: $(1,4,5)(2,3)(7,8) = (7,8)(2,3)(1,4,5)$.

□

Corollary 15. *If $\sigma_1, \sigma_2, \cdots, \sigma_k$ are disjoint cyclic permutations, each with respective orders of n_i, then the order of $\sigma_1\sigma_2 \cdots \sigma_k$ is $\operatorname{lcm}(n_1, n_2, \cdots, n_k)$*

Proof.
Let n_i be the order of σ_i for all i such that $1 \leq i \leq k$. Let r denote the order of $\sigma_1\sigma_2 \cdots \sigma_k$ and $l = \operatorname{lcm}(n_1, n_2, \cdots, n_k)$. Since the cycles are disjoint, they commute and so

$$(\sigma_1\sigma_2 \cdots \sigma_k)^l = \sigma_1^l\sigma_2^l \cdots \sigma_k^l = e.$$

Thus r divides l. Since r is the order of $\sigma_1\sigma_2\cdots\sigma_k$ we know $(\sigma_1\sigma_2\cdots\sigma_k)^r = e$. This implies $\sigma_i^r = e$ for all $1 \leq i \leq k$. Thus r divides n_i for each i and r divides $\text{lcm}(n_1, n_2, \cdots, n_k)$.

□

Definition:

A cycle of length two is a *transposition*.

□

Corollary 16. *Let $\sigma \in S_n$ for $n \geq 2$. This permutation σ is a product of transpositions.*

Proof.
Every permutation can be written in terms of disjoint cycles, so we only need to show that every cycle can be written as a product of transpositions. We see that $(a_1, a_2, a_3) = (a_1, a_3)(a_1, a_2)$ and in general for cycle of length k:

$$(a_1, a_2, \cdots, a_{k-1}, a_k) = (a_1, a_k)(a_1, a_{k-1})\cdots(a_1, a_3)(a_1, a_2).$$

□

Example: Consider the effect of the cycle $(1, 5, 4, 3)$ on the finite set $\{1, 2, 3, 4, 5\}$. We can rewrite this cycle as a product of transpositions: $(1, 5, 4, 3) = (1, 3)(1, 4)(1, 5)$:

$$1\ 2\ 3\ 4\ 5 \overset{(1,5)}{\rightarrow} 5\ 2\ 3\ 4\ 1 \overset{(1,4)}{\rightarrow}$$

$$5\ 2\ 3\ 1\ 4 \overset{(1,3)}{\rightarrow} 5\ 2\ 1\ 3\ 4$$

□

Definition:
A permutation of a finite set is *even* if it can be expressed as an even number of transpositions. A permutation is *odd* if it can be expressed as an odd number of transpositions.

□

There is nothing that we have said that excludes the possibility that a permutation could be written as both an even and an odd number of transpositions. We will see that there is no way for a permutation to be both even and odd.

Theorem 17.
If a permutation σ is expressed in terms of m transpositions and in another way as k transpositions with $k < m$, then m and k are both even or both odd.

Proof.
We begin by stating that the identity ι can only be even. Begin with the identity, if we apply a transposition τ to ι then we must also apply τ^{-1}. It follows that *i* is even.

Let σ be a permutation which is expressed in terms of m transpositions τ_i and k transpositions μ_i. We have $\tau_1\tau_2\cdots\tau_m = \mu_1\mu_2\cdots\mu_k$. Thus $\tau_1\tau_2\cdots\tau_m\mu_k\cdots\mu_2\mu_1 = i$. This means that $\tau_1\tau_2\cdots\tau_m\mu_k\cdots\mu_2\mu_1$ is even. Thus $m+k$ is an even number, so either both m, k are even, or m, k both odd.

□

This definition and subsequent theorem give us a way to classify any permutation as either even or odd. The number of transposition used to express a permutation may vary. For an easy example of this, consider the identity permutation

$$i = (1,2)(1,2) = (1,2)(3,4)(1,2)(3,4).$$

There is a different number of transpositions for each of these, but in all cases there is an even number of them.

We can compose permutations as we have been doing. It is worthwhile to consider what happens when both τ and σ are even permutations. The result of composing these is $\tau\sigma$. Due to the fact that the sum of two even numbers is even, we see that $\tau\sigma$ is an even permutation. This means that we can consider the composition of two permutations a binary operation on the set of *even* permutations. This with the next theorem explains why we care about classifying permutations as even or odd.

Theorem 18.
If $n \geq 2$ then the set of all even permutations of $\{1, 2, 3, \cdots, n\}$ with binary operation composition of permutations forms a group.

Proof.
We must check that the conditions of a group are satisfied:

- We have already seen that the composition of any permutation is associative. It follows that for any even permutations σ, τ, μ we have
$$\sigma \circ (\tau \circ \mu) = (\sigma \circ \tau) \circ \mu.$$
- The identity permutation ι is even, since $(1,2)(2,1) = \iota$ is the identity.
- If σ is even then it can be written as an even number of transpositions $\sigma = \tau_1\tau_2\cdots\tau_{2k}$. We may express $\sigma^{-1} = \tau_{2k}\cdots\tau_2\tau_1$, so σ^{-1} is also even.

□

Although it would seem obvious that exactly half of the permutations of S_n are even and half are odd, it is worthwhile to carefully prove this statement. Sometimes statements that seem to be intuitively obvious turn out to not be the case. In this case it is true that permutations are split exactly in half between even and odd permutations.

Theorem 19.
For $n \geq 2$ There are $n!/2$ even permutations in S_n.

Proof.
Let E_n denote the set of even permutations of S_n, thereby making E_n^C the set of odd permutations. Since $n \geq 2$ we know that at least one transposition τ exists. Define a mapping $f : E_n \to E_n^C$ where $f(\sigma) = \tau\sigma$. As we can see from this definition, this maps an even permutation to an odd permutation by composing an additional transposition to a permutation. We now show that the number of elements in E_n and E_n^C are equal by showing that f is a one-to-one function.

Suppose that $f(\sigma_1) = f(\sigma_2)$. Thus $\tau\sigma_1 = \tau\sigma_2$. We apply the inverse of τ (which is τ itself) on the left of both sides of the equation and see that $\tau\tau\sigma_1 = \tau\tau\sigma_2 \Rightarrow \sigma_1 = \sigma_2$. Therefore the mapping f is one-to-one and the number of elements in E_n is the same as the number of elements in E_n^C. This is exactly half of the total number of permutations of a set of n elements, and so there are $n!/2$ even permutations.

□

Definition:
The *alternating group on* n *elements* A_n is the set of all even permutations of the set $\{1, 2, \cdots, n\}$. The order of A_n is $n!/2$

□

It turns out that alternating groups provide us with a very important class of groups to study. These groups are connected to the symmetric groups that we have studied. What is perhaps more useful is that a certain property of A_n for $n \geq 5$ is crucial to proving that certain polynomial equations of degree five or greater do not have a solution expressed in terms of radicals and basic arithmetic.

2.6.1 Exercises

1. For each of the following, write the orbits, write cycle notation for each permutation, and determine if σ an even or odd permutation:
   a) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 2 & 5 & 1 & 6 \end{pmatrix}$
   b) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 1 & 2 & 6 & 3 \end{pmatrix}$
   c) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 4 & 5 & 6 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 2 & 4 & 3 & 1 \end{pmatrix}$
2. Is there an element of order 6 in A_5?
3. Compute the product of cycles $(1, 4, 5, 6)(4, 3, 2, 1)$
4. Express the permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 1 & 4 & 6 & 3 \end{pmatrix}$
   a) As a product of disjoint permutations
   b) As a product of transpositions.
5. Show that every even permutation of three or more elements can be written as the product of cycles of length 3.
6. Let H be a subgroup of S_n for $n \geq 2$. Show that H is a subgroup of A_n or exactly half of the elements of *H* are even, half are odd.

2.7 Subgroups

We have examined several different kinds of groups. Each of these were defined by a different property. Indeed, there is a world a difference between the description of the cyclic group $\mathbb{Z}_{12}$ and the alternating group A_4. Beyond considerations of the definition of the group and the total number of elements of the group, there are some other features that we can explore. One of these is the internal structure of the group.

As we have seen with the alternating group A_n in relation to the symmetric group S_n, sometimes a subset of elements of a group is a group in its own right. We wish to name this feature and study aspects of it.

Definition:
Let $\cdot$ be a binary operation on *S* and *T* a subset of *S*. If for all $a, b \in T$, $a \cdot b \in T$ we say that *T* is *closed under* $\cdot$

□

Example:
The following are a series of example and counterexamples of closure under a specified binary operation.

- The subset $\mathbb{Z}$ of $\mathbb{R}$ is closed under the binary operation of standard addition.
- The subset $\mathbb{R}^*$ – the nonzero real numbers – is not closed under standard addition because $3, -3 \in \mathbb{R}^*, \text{ but } 3 + (-3) = 0 \notin \mathbb{R}^*$.
- Even integers are closed under standard addition and multiplication. This is really just a more formal way of sayings "even plus even is even" and "even times even is even."
- The odd integers are closed under standard multiplication. This is a more formal way of saying "odd times odd is odd."
- The odd integers are not closed under standard addition. To show this we have $3 + 5 = 8$. In more generality, we can add $(2k + 1), (2n + 1)$ where $n, k \in$. We see that $(2k + 1) + (2n + 1) = 2(n + k + 1)$, an even number.
- We have seen that the even permutations are closed under the binary operation of composition of permutations.

□

Definition:

If a subset *H* of a group is closed under the binary operation of *G* and if *H* is a group, then *H* is a *subgroup* of *G*, and is denoted $H \leq G$.

□

There are two subgroups that every group possesses. Since a set is a subset of itself, the group *G* is a subgroup of itself. The other subgroup that every group has is the subgroup consisting of only the identity element. Other subgroups can be varied, and more interesting than these two. This gives us the following definitions:

Definition:

The *trivial subgroup* of the group *G* is the identity element alone.

□

Definition:

A *proper subgroup* of the group *G* is any subgroup other than the group *G* itself.

□

Example:
We will examine a series of examples and counterexamples of subgroups.

1. $\mathbb{Z}$ under standard addition is a subgroup of $\mathbb{R}$ under standard addition.
2. $\mathbb{Q}^* \subset \mathbb{R}$ and $\mathbb{Q}^*$ under standard addition is a group, however this is not a subgroup of $\mathbb{R}$ with binary operation addition. The reason why is that the operations do not match.
3. We let C_n denote the group of complex n th roots of unity. Each of these groups is a subgroup of the group $\mathbb{C}^*$ under multiplication of complex numbers.
4. A_n is a subgroup of S_n for $n \geq 1$.
5. The group $\mathbb{Z}_4$ has the trivial subgroup $\{0\}$, improper subgroup $\mathbb{Z}_4$, and subgroup $\{0, 2\}$.
6. The Klein four group V has the trivial subgroup $\{e\}$, improper subgroup V, and three other subgroups $\{e, a\}, \{e, b\}, \{e, c\}$.

□

Theorem 20. *A subset H of a group G is a subgroup if and only if*

1. *H is closed under the binary operation of G*
2. *The identity element e of G is also in H*
3. *For all $a \in H$ $a^{-1} \in H$ also.*

Proof.
If *H* is a subgroup of *G* then it is a group under the binary operation $\cdot$ of *G*. Thus *H* is closed. *H* must contain the identity element. If $a \in H$ then $a^{-1} \in H$ since *H* is a group.

Now suppose that the list of conditions hold. This shows that *H* is a group as it inherits the associative structure from *G*. It follows that *H* is a subgroup of *G*.

□

The advantage to knowing this theorem is that it makes it easier to show that a subset of a group is a subgroup. If any one of the conditions in the list does not hold, then we automatically know that *H* is not a subgroup.

Example:

Show that the set of odd integers *H* does not form a subgroup of $\mathbb{Z}$ under addition.

This follows very quickly from the fact that the identity element $0 \in \mathbb{Z}$ in not odd, and so $0 \notin H$. We could also argue that this is not a subgroup because *H* is not closed under the binary operation.

□

Example: Now consider $\mathbb{Z}_8$ under addition modulo 8. If *H* is a subgroup of $\mathbb{Z}_8$ that contains 2 but not 1 then what else do we know about this subgroup?

By the theorem it is immediate that 0 (the identity) and 6 (the inverse of 2) are also elements in *H*. By the closure property $2 + 2 = 4 \in H$. Thus $H = \{0, 2, 4, 6\}$.

□

The subgroup in the previous example is known as a cyclic subgroup. This is because it was generated by a single element 2, just like our cyclic groups were.

Definition: Let *G* be a group and let $a \in G$ then the subgroup $\{a^n | n \in \mathbb{Z}\}$ is called the *cyclic subgroup* of *H* generated by a and is denoted $\langle a \rangle$.

□

2.7.1 Some Number Theory

At this point we note some results from the area of mathematics known as number theory. These facts are interesting for their own sake, and will be needed in some of the proofs that follow.

Definition: Let r and s be two positive integers. The *greatest common divisor* of r and s, denoted $d = \gcd(r, s)$ is the greatest integer that divides both r and s.

□

Definition: Two positive integers are *relatively prime* if their gcd is 1

□

Definition: The least common multiple of r and s, denoted $\text{lcm}(r, s)$, is the least positive integer that is divisible by both r and s.

□

Note: Let $d = \gcd(r, s)$ then $d \cdot \text{lcm}(r, s) = rs$

□

We will now formalize the process of long division. We will show that we may divide any integer by a positive integer and obtain a unique quotient q and remainder r.

Theorem 21 [Division Algorithm] *Let $n, p \in \mathbb{Z}$ with $p > 0$. Then there exist unique $p, r \in \mathbb{Z}$ such that $n = pq + r$ and $0 \leq r < p$.*

Proof. Let $S = \{x \in \mathbb{Z}^+ \mid x = n - pm, m \in \mathbb{Z}\}$.. We let r be the smallest number in the set S. There is some $q \in \mathbb{Z}$ such that $r = n - pq$ and $r < p$. If $r \geq p$ the $r - p > 0$ and $r - p = n - pq - p = n - p(q + 1) \in S$, contradicting that r is the smallest positive integer in the set S. Uniqueness of r is automatic, and uniqueness of q follows from if $r = n - pq$ and $r = n - pq'$ then $q = q'$.

□

Example: Use the division algorithm for $25 \div 7$

$25 = 3 \cdot 7 + 4$

□

We see that this is just an alternate way of expressing a long division problem. Besides being used for division, the division algorithm has other uses. One that is very helpful it that it can be used repeatedly to find the greatest common divisor of two positive integers.

Example:

Find $\gcd(24, 138)$ by using the division algorithm. 24 and 138:

$138 = 24 \cdot 5 + 18$

$24 = 18 \cdot 1 + 6$

$$18 = 6 \cdot 3 + 0$$

The last nonzero remainder is the greatest common divisor of the numbers we started the process with.

□

2.7.2 Subgroup Theorems

Now that our brief excursion to number theory is over, we can use these ideas to prove statements regarding subgroups.

Theorem 22. *Every subgroup of a cyclic group is cyclic.*

Proof. Let a be a generator of the cyclic group G, and suppose that H is a subgroup of G. If $H = \{e\}$ then we are done, as the subgroup $\{e\}$ is a cyclic group.

If $H \neq \{e\}$ then there is some element $b \in H$ such that $b \neq e$. Since a is a generator of G, and $b \in G$, there exists some $n \in^+$ such that $a^n = b$. So $a^n \in H$. Since this is true for some positive integer, there is a least positive integer for which this statement is true. Let k be the least positive integer k such that $a^k \in H$. The goal is to show that a^k is a generator of H.

Let $c \in H$. Since H is a subgroup of cyclic group G, $c = a^m$ for some m. By the division algorithm there exist q and r with $0 \leq r < k$ such that $m = kq + r$:

$$a^m = a^{kq+r} = (a^k)^q a^r \Rightarrow a^r = a^m (a^k)^{-q} = c(a^k)^{-q}.$$

Since $c, a^k \in H$, This shows that $a^r \in H$. Since a^r is in H, $0 \leq r < k$, and since k is the least positive integer such that $a^k \in H$, it follows that r is not positive and $r = 0$. Therefore $m = kq$ and $a^m = (a^k)^q$ thus a^k generates H, making H cyclic.

□

Theorem 23. *A nonempty subset H of the group G is a subgroup of G if and only if for all $a, b \in H$, $a^{-1}b \in H$.*

Proof. Begin by supposing that H is a subgroup of G. For $a, b \in H$, it follows that $a^{-1} \in H$ and $a^{-1}b \in H$ by the closure of H.

Now suppose that H is a nonempty subset where for all $a, b \in H$, $a^{-1}b \in H$. From this $e = a^{-1}a \in H$. Since the identity element $\{e\} \in H$ we also know that $a^{-1} = a^{-1}e \in H$. Also, for every $a, b \in H$ the element $\left(a^{-1}\right)^{-1} b = ab \in H$ so H is closed.

□

Theorem 24. If H and K are both subgroups of the group G then $H \cap K$ is also a subgroup of G.

Proof.
The set $H \cap K$ is closed because if $a, b \in H \cap K$ then $a, b \in H$ and $a, b \in K$. Since H and K are both closed, $ab \in H$ and $ab \in K$. Thus $ab \in H \cap K$.

The identity element $\{e\} \in H$ and $\{e\} \in K$, so $\{e\} \in H \cap K$.

If $a \in H \cap K$ then $a \in H$ and $a \in K$. Since H and K are subgroups, $a^{-1} \in H$ and $a^{-1} \in K$. By the above we know that $H \cap K$ is a subgroup of G.

□

2.7.3 Exercises

1. Let H consists of the elements of the group G such that $x \cdot x = e$. Show that H is a subgroup of G
2. List all subgroups of D_3.
3. List all subgroups of D_4.
4. List all subgroups of $\mathbb{Z}_{36}$

5. List all subgroups of $\mathbb{Z}_{60}$
6. List all subgroups of A_4.

2.8 Homomorphisms and Isomorphisms

We are interested in studying the different types of structures that binary operations can provide on sets with the same number of elements:

$\cdot$	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Note that the tables are the same if we replace as follows: $a \leftrightarrow 0 \quad b \leftrightarrow 1 \quad c \leftrightarrow 2$. In other words, there is a one-to-one, onto mapping between $\{a, b, c\}$ and $\{0, 1, 2\}$. There is actually one more feature that this mapping possesses. This will be the first topic that we will address in this section.

2.8.1 Homomorphisms

The extra feature that we need for our mapping is that it satisfies the homomorphism property. This is needed in order for the binary operations of both groups to match.

Definition: A map ϕ of a group G with binary operation $\cdot$ into a group G' with binary operation $\#$ is a *homomorphism* if $\phi(a \cdot b) = \phi(a) \# \phi(b)$ holds for all $a, b \in G$.

Show that the following $\phi(\sigma) = \begin{cases} 0 & \text{if} \quad \sigma \text{is an even permutation} \\ 1 & \text{if} \quad \sigma \text{is an odd permutation} \end{cases}$ is a homomorphism $\phi : S_n \to_2$ □

We check that this is a homomorphism by looking at the four cases that are possible:

- If σ and τ are even, then $\sigma\tau$ is even:

$$0 = \phi(\sigma\tau) = \phi(\sigma) + \phi(\tau) = 0 + 0.$$

- If σ is odd and τ is even, then $\sigma\tau$ is odd:

$$1 = \phi(\sigma\tau) = \phi(\sigma) + \phi(\tau) = 1 + 0.$$

- If σ is even and τ is odd, then $\sigma\tau$ is odd:

$$1 = \phi(\sigma\tau) = \phi(\sigma) + \phi(\tau) = 0 + 1.$$

- If σ and τ are odd, then $\sigma\tau$ is even:

$$0 = \phi(\sigma\tau) = \phi(\sigma) + \phi(\tau) = 1 + 1.$$

□

Since a homomorphism is a particular kind of mapping, not every mapping from one group to another is a homomorphism. There is at least one homomorphism mapping one group to another, but it is not very complicated (or interesting).

Definition: The trivial homomorphism $\phi : G \to G$ is $\phi(g) = e'$ for all $g \in G$.

□

Example:

The mapping $\phi : \mathbb{Z} \to 7\mathbb{Z}$ given by $\phi(x) = 7x$ is a homomorphism of $\mathbb{Z}$ onto the set of multiplies of 7.

Let $x, y \in \mathbb{Z}$. We see that $\phi(x + y) = 7(x + y) = 7x + 7y = \phi(x) + \phi(y)$ and so ϕ is a homomorphism.

□

Example:

The mapping $\phi : \mathbb{Z}_4 \to \mathbb{Z}_4$ given by $\phi(1) = 3$ is a homomorpism. We see that this implies:

- $\phi(2) = \phi(1 + 1) = \phi(1) + \phi(1) = 3 + 3 = 2$
- $\phi(3) = \phi(2 + 1) = \phi(2) + \phi(1) = 2 + 3 = 1$
- $\phi(0) = \phi(2 + 2) = \phi(2) + \phi(2) = 2 + 2 = 0$

□

Theorem 25. *Given group G with binary operation $\cdot$ and G' with binary operation $\#$, if $\phi : G \to G'$ is a group homomorphism and $e \in G$ is the identity, then $\phi(e)$ is the identity of G'.*

Proof.
Since ϕ is a homomorphism we know that $\phi(x \cdot y) = \phi(x)\#\phi(y)$ for all $x, y \in G$. Since e is the identity of *G* we have:

$$\phi(x)\#\phi(e) = \phi(x \cdot e) = \phi(x) = \phi(e \cdot x) = \phi(e)\#\phi(x)$$

and so by definition $\phi(e)$ is the identitity of G'.

□

Theorem 26. *Given the group homomorphism $\phi : G \to G$, for any $g \in G, [\phi(g)]^{-1} = \phi(g^{-1})$.*

Proof. For $g \in G$ consider the element $\phi(g)$. Since ϕ is a homomorphism:

$$\phi(g^{-1})\#\phi(g) = \phi(g^{-1}g) = \phi(gg^{-1}) = \phi(g)\#\phi(g^{-1}).$$

Furthermore, since $\phi(gg^{-1}) = \phi(e) = e$ we have $\phi(g^{-1})\#\phi(g) = \phi(g)\#\phi(g^{-1}) = e$ and so $[\phi(g)]^{-1} = \phi(g^{-1})$.

□

We have seen that homomorphisms are mappings that take the identity of *G* to the identity of *G*. Also, a homomorphism maps inverses of *G* to inverses in *G*. The subgroup structures of *G* and *G* are also mapped to each other by homomorphisms.

Definition: Let ϕ be a mapping of a set X into a set Y, and let $A \subseteq X$ and $B \subseteq Y$.

- The *direct image* of A is the set $\phi(A) = \{\phi(a) | a \in A\}$.
- The *inverse image* of B is the set $\phi^{-1}(B) = \{x \in X | \phi(x) \in B\}$.

□

Theorem 27. *Let ϕ be a homomorphism of a group G into a group G'. If H is a subgroup of G, then $\phi[H]$ is a subgroup of G'.*

Proof. Let H be a subgroup of the group *G* and ϕ a homomorphism from *G* to *G*. We begin by showing that the set $\phi(H)$ is closed in *G*. If $y_1, y_2 \in \phi(H)$, then there exist $h_1, h_2 \in H$ such that $y_1 = \phi(h_1)$ and $y_2 = \phi(h_2)$. Thus $y_1 y_2 = \phi(h_1)\phi(h_2) = \phi(h_1 h_2) \in \phi(H)$.

Since $e \in H$, $\phi(e) = e \in \phi(H)$. So $\phi(H)$ has the identity element.

For $y \in \phi(H)$ there exists $h \in H$ such that $y = \phi(h)$. Since H is a subgroup, $h^{-1} \in H$. We have $\phi(h^{-1}) = [\phi(h)]^{-1} = y^{-1} \in \phi(H)$. Therefore $\phi(H)$ is a subgroup of *G*.

□

Example:

We consider the homomorphism $\phi : \mathbb{Z}_{12} \to \mathbb{Z}_{12}$ defined by $\phi(1) = 3$, and look at the images of all of the subgroups of $\mathbb{Z}_{12}$.

- $\phi(\langle 0\rangle) = \langle 0\rangle$
- $\phi(\langle 1\rangle) = \phi(\{0,1,2,3,4,5,6,7,8,9,10,11\}) = \{0,3,6,9\} = \langle 3\rangle$
- $\phi(\langle 2\rangle) = \phi(\{0,2,4,6,8,10\}) = \{0,6\} = \langle 6\rangle$
- $\phi(\langle 3\rangle) = \phi(\{0,3,6,9\}) = \{0,3,6,9\} = \langle 3\rangle$
- $\phi(\langle 4\rangle) = \phi(\{0,4,8\}) = \{0\} = \langle 0\rangle$
- $\phi(\langle 6\rangle) = \phi(\{0,6\}) = \{0,6\} = \langle 6\rangle$

We note that the subgroup inclusions are respected by the homomorphism. For instance, just as $\langle 4\rangle$ is a subgroup of $\langle 2\rangle$, $\phi(\langle 4\rangle)$ is a subgroup of $\phi(\langle 2\rangle)$.

□

Theorem 28. *Let ϕ be a homomorphism of a group G into a group G'. If K is a subgroup of G', then $\phi^{-1}[K]$ is a subgroup of G.*

Definition: Let $\phi : G \to G'$ be a homomorphism of groups. Then the subgroup

$$\phi^{-1}(\{e\}) = \{x \in G | \phi(x) = e'\}$$

is the *kernel* of ϕ, denoted $\ker\phi$.

□

Note: Since e is a subgroup of any group, this is a particular instance showing that the inverse image of a subgroup is a subgroup.

□

Theorem 29. *A group homomorphism $\phi : G \to G'$ is a one-to-one map if and only if* $\ker\phi = \{e\}$.

Proof. Suppose that $\ker\phi = \{e\}$. If $\phi(g_1) = \phi(g_2)$ then it follows that

$$\phi(g_1)[\phi(g_2)]^{-1} = e \Rightarrow \phi(g_1 g_2^{-1}) = e.$$

This implies that $g_1 g_2^{-1} \in \ker \phi$. As $\ker \phi$ has only one element, $g_1 g_2^{-1} = e$ and so $g_1 = g_2$. The mapping ϕ is one-to-one.

Now suppose that ϕ is one-to-one. We know that homomorphisms map the identity element of G to the identity element of G. In other words, $\phi(e) = e'$. Since ϕ is one-to-one, this is the only element of G mapped into e' by ϕ. Therefore $\ker\phi = e$.

□

2.8.2 Isomorphisms

We return to our original question. How do we show that the underlying structures of two groups are the same? It's clear that both of the groups being considered must have the same number of elements. So there must be a one-to-one onto mapping between the groups. The additional property that such a mapping must possess is that it be a homomorphism. This will ensure that the binary operations of the two groups match.

Definition: Let G with binary operation $\cdot$ and G with binary operation $\#$. An *isomorphism of G with G* is a one-to-one onto homomorphism $\phi : G \to G$. For all $x, y \in G$:

$$\phi(x \cdot y) = \phi(x) \# \phi(y)$$

If such a mapping exists, we say G is isomorphic to G and write $G \cong G$.

□

Example:

Let $\mathbb{Z}_5$ denote equivalence classes modulo 5 with binary operation addition and C_5 the complex fifth roots of unity with binary operation of complex multiplication. Set any primitive fifth root of unity equal to ξ and define $\phi : C_5 \to_5$ by $\phi(\xi) = 1$.

The mapping ϕ is one-to-one. Suppose $\phi(x) = \phi(y) = n \in \_5$. Thus $x = \xi^{5k+n}$ and $y = \xi^{5m+n}$. Thus $x = \xi^{5k+n} = \xi^n = \xi^{5m+n} = y$ Thus ϕ is one-to-one.

ϕ is onto since for $0 \leq n \leq 4$, $\phi(\xi^n) = n$.

By the following we have an isomorphism:

$$\phi(x \cdot y) = \phi(\xi^{5k+i} \cdot \xi^{5m+j}) = \phi(\xi^{i+j}) =$$
$$i+j \bmod 5 = i \bmod 5 + j \bmod 5 = \phi(\xi^i) + \phi(\xi^j)\phi(\xi^{5k+i}) + \phi(\xi^{5m+j}) = \phi(x) + \phi(y)$$

□

To demonstrate that two groups do not have isomorphic structures, we must examine structural properties of the groups in question. If we can demonstrate that any of these properties are different, then we can state definitively that the groups are not isomorphic.

Example:

1. The groups $\mathbb{Z}_2$ and V are not isomorphic, because they are not of the same order.
2. The groups $\mathbb{Z}_4$ and V are not isomorphic. Even though they are both of order four, one is cyclic and the other is not.
3. The groups $\mathbb{Z}_6$ and D_3 are not isomorphic. Even though they are both of order six, one is abelian and the other is not.
4. $(\mathbb{C}, \cdot), (\mathbb{R}, \cdot)$ are not isomorphic. $x \cdot x = c$ always has a solution for x in $\mathbb{C}$ but $x \cdot x = -1$ does not have a solution in $\mathbb{R}$.

□

The following theorem explains why we were so interested in studying S_n and groups of permutations.

Theorem 30 (Cayley's Theorem) *Every finite group of order n is isomorphic to a group of permutations of a set with n elements.*

2.8.3 Cyclic Group Structure

It turns out that up to isomorphism cyclic groups have a relatively straightforward description.

Theorem 31. *Let G be a cyclic group with generator a. If the order of G is infinite, then G is isomorphic to $\mathbb{Z}$ with binary operation of addition. If G has finite order n then G is isomorphic to $\mathbb{Z}_n$ with binary operation of addition modulo n.*

Proof. We split this proof into two cases, which will correspond to the infinite cyclic groups and finite cyclic groups. We first consider the possibility that for all integers $k > 0$ we have $a^k \neq e$. Suppose that $a^m = a^j$ and that $j > m$. Then $a^j a^{-m} = a^{j-m} = e$, which is a contradiction. Thus every element of *G* can be uniquely expressed as a^k for $k \in \mathbb{Z}^+$. Define the map $\phi : G \to \mathbb{Z}$ by $\phi(a^k) = k$. This mapping is well-defined, one-to-one, and onto.

$$\phi(a^j a^m) = \phi(a^{j+m}) = j + m = \phi(a^j) + \phi(a^m)$$

Therefore ϕ is an isomorphism between this cyclic group and the integers $\mathbb{Z}$.

Now we suppose that there exists a positive integer k such that $a^k = e$. Let n be the smallest such positive integer. If $u \in$ and $u = nq + r$ for $0 \leq r < n$, then $a^u = a^{nq+r} = (a^n)^q a^r = a^r$. By a similar argument as the previous case, the elements $e = a^0, a, a^2, \cdots, a^{n-1}$ are all distinct and comprise all of *G* the map $\phi : G \to_n$ given by $\phi(a^j) = j$ is well defined, one to one, and onto. We see that

$$\phi(a^j a^m) = \phi(a^{j+m}) = (j + m) \bmod n = j \bmod n + m \bmod n = \phi(a^j) + \phi(a^m)$$

therefore ϕ is an isomorphism between $\mathbb{Z}_n$ and the cyclic group *G*.

□

2.8.4 Exercises

1. Let $\phi : G \to G'$ be a homomorphism of *G* onto G'. Prove that if *G* is abelian, then G' is abelian.
2. Show that group isomorphism is an equivalence relation on groups.
3. Consider the se $\mathbb{Z}$ under standard addition. Is the mapping $\phi : \mathbb{Z} \to \mathbb{Z}$ defined by $\phi(n) = -n$ an isomorphism?
4. Let G be the set of matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $a, b \in$.
   a) Prove that *G* is closed under matrix addition and matrix multiplication.
   b) Prove that *G* with matrix addition is isomorphic to $\mathbb{C}$ with addition of complex numbers.
   c) Prove that *G* with matrix multiplication is isomorphic to $\mathbb{C}$ with multiplication of complex numbers.
5. An *automorphism* is an isomorphism of a group with itself.
   a) How many automorphisms does $\mathbb{Z}_{10}$ have?
   b) How many autormorphisms does $\mathbb{Z}_7$ have?
   c) How many automorphisms does $\mathbb{Z}_p$ for p prime have?

6. Let $\phi : \mathbb{Z}_{12} \to \mathbb{Z}_4$ be defined by $\phi(x) = x \bmod 4$. Prove that this is a homomorphism.
7. Let $\phi : \mathbb{Z} \to \mathbb{Z}_{13}$ be defined by $\phi(1) = 7$. Determine $\ker\phi$ and $\phi(57)$.
8. Let $\phi : \mathbb{Z} \to \mathbb{Z}$ be a homomorphism
   a) Determine the homomorphisms ϕ that are onto.
   b) Determine the total number of homomorphisms ϕ.
9. Let $\phi : G \to G'$ be a group homomorphism. Prove that if the order of G is prime, then ϕ is trivial or one-to-one.

2.9 Cosets and Normal Subgroups

In our examination of subgroups we saw some examples that seemed to suggest if H is a subgroup of G then the order of n divides the order of k. We pause for the reminder that in mathematics several examples are not enough to prove a theorem, but a single counterexample is enough to disprove a statement. In this case we are okay, as the order of a subgroup does in fact divide the order of the group that it is contained in. In this section we will prove that this is true.

2.9.1 Cosets

Definition:

Let H be a subgroup of a group G. Define the relation $\prec$ by $x \prec y$ if and only if $x^{-1}y \in H$.

□

Theorem 32. *The relation defined above is an equivalence relation.*

Proof. We check that the three conditions of an equivalence relation are met:

- Reflexive: Since H is a subgroup, $g^{-1}g = e \in H$ for all $g \in G$. Thus $g \prec g$ for all $g \in G$
- Symmetric: Suppose that $x \prec y$. Thus $x^{-1}y \in H$. Since H is a subgroup $(x^{-1}y)^{-1} = y^{-1}x \in H$. This means that $y \prec x$.
- Transitive: Suppose that $x \prec y$ and $y \prec z$. So we have $x^{-1}y \in H$ and $y^{-1}z \in H$. By closure of H we know $(x^{-1}y)(y^{-1}z) = x^{-1}z \in H$. So $x \prec z$.

□

The equivalence classes of this equivalence relation are the following types of sets.

Definition: Let H be a subgroup of a group *G*. The subset $gH = \{gh \mid h \in H\}$ of *G* is the *left coset* of H containing g, while $Hg = \{hg \mid h \in H\}$ is the *right coset* of H containing g.

□

Example: $4\mathbb{Z} = \{\cdots, -12, -8, -4, 0, 4, 8, 12, \cdots\}$ is a subgroup of $\mathbb{Z}$. The cosets are:

- $4\mathbb{Z} = \{\cdots, -12, -8, -4, 0, 4, 8, 12, \cdots\}$
- $1 + 4\mathbb{Z} = \{\cdots, -11, -7, -3, 1, 5, 9, 13, \cdots\},$, which is obtained by adding 1 to each of the elements of $4\mathbb{Z}$.
- $2 + 4 = \{\cdots, -10, -6, -2, 2, 6, 10, 14, \cdots\}$ which is obtained by adding 2 to each of the elements of $4\mathbb{Z}$.
- $3 + 4 = \{\cdots, -9, -5, -1, 3, 7, 11, 15, \cdots\}$ which is obtained by adding 3 to each of the elements of $4\mathbb{Z}$.

□

Example:

$\langle 4 \rangle = \{0, 4, 8\}$ is a subgroup of $\mathbb{Z}_{12}$.

$\{1, 5, 9\}$ is the coset containing 1, $\{2, 6, 10\}$ is the coset containing 2, $\{3, 7, 11\}$ is the coset containing 3, and $\{0, 4, 8\}$ is the coset containing 0.

□

Example:

$\langle 3 \rangle = \{0, 3, 6, 9\}$ is a different subgroup of $\mathbb{Z}_{12}$.

$\{1, 4, 7, 10\}$ is the coset containing 1, $\{2, 5, 8, 11\}$ is the coset containing 2, and $\{0, 3, 6, 9\}$ is the coset containing 0.

□

Note:

For any group *G* and subgroup *H*, the coset of *H* containing the identity is also a subgroup of *G*. However, no other cosets are subgroups. (none of the other cosets contain the identity). Cosets are *subsets* of the group *G*.

□

Definition:

Let *H* be a subgroup of the group *G*. The number of distinct left (right) cosets of *H* is called the *index of H in G* and is denoted $(G : H)$.

□

We have seen that cosets partition our group into several disjoint sets. It turns out that these disjoint sets all have the same number of elements. After carefully proving this result, which does not appear that interesting on the surface, we will see that this equal partitioning implies a significant result about the order subgroups of a group.

Theorem 33. *Let H be a subgroup of G and* g *any element of G. The coset* gH *has the same number of elements as H.*

Proof. Define a map $\phi : H \to gH$ by $\phi(h) = gh$ for all $h \in H$. We will see that ϕ is one-to-one:

$$\phi(h_1) = \phi(h_2) \Rightarrow gh_1 = gh_2 \Rightarrow h_1 = h_2$$

Since ϕ is one-to-one, we see that the sets *H* and gH have the same number of elements. For *H* an infinite set this implies that gH is also infinite.

□

Although we have been working with left cosets, there is no reason why all of the above discussion couldn't have focused on right cosets. It is also true that *H* and Hg have the same number of elements.

Theorem 34 (Lagrange's Theorem). *Let H be a subgroup of a finite group G. Then the order of H is a divisor of the order of G.*

Proof. Let n be the order of G and H have order m. Every coset of H also has m elements. Let r be the number of distinct left cosets of H. By considering the group G as $G = H \cup g_1H \cup g_2H \cup \cdots \cup g_{r-1}H$, we have the equation $n = rm$. Therefore m is a divisor of n.

□

WARNING: The converse of Lagrange's theorem is not true. Just because k divides the order of a group n does not mean that group has a subgroup of order k. A specific class of counterexamples will be seen when we study direct products.

□

Corollary 35. *Every group with order a prime number is a cyclic group.*

Let G be of prime order p. Since G has prime order there are at least two elements. Let $a \in G$ be an element different from the identity. Then $\langle a \rangle$ has at least two elements $\{e, a\}$. By Lagrange's Theorem, the order n of $\langle a \rangle$ must divide the order of G. In other words, n must divide p. As $n > 1$, it is immediate that $n = p$. Therefore $G = \langle a \rangle$.

□

This corollary is quite powerful as it tells us that up to isomorphism, there is only one group of order $\mathbb{Z}\, p$. Since this group is cyclic, it is isomorphic to $\mathbb{Z}_p$ and is abelian. We will soon see that this is quite an accomplishment to be able to make such a definitive statement. In general it is quite difficult to tell how many groups exist of a particular order.

2.9.2 Normal Subgroups

When we defined the cosets of H containing g we made a distinction between the left and the right cosets. There was a good reason for this. Due to the fact that some groups are nonabelian, in general $gH \neq Hg$. It will become apparent that it is worthwhile to distinguish between subgroups H of a group G for which $gH = Hg$ for all $g \in G$, and those for which this property is not true. Worthwhile enough, in fact, for a definition.

Definition: A subgroup H of G is a *normal subgroup* if $gH = Hg$ for all $g \in G$.

□

Theorem 36. *If G is abelian and H is a subgroup of G then H is normal.*

Proof.
The proof is immediate from the fact that we have an abelian group. If $H = \{h_1, h_2, \cdot\}$ we see that

$$gH = \{gh_1, gh_2, \cdots\} = \{h_1 g, h_2 g, \cdots\} = Hg.$$

□

Although left and right cosets match for abelian groups, for nonabelian groups this is not the case. Left and right cosets in nonabelian groups may or may not coincide with one another.

Example:

We consider the dihedral group D_3 and consider the cosets of two of this groupï¿½s subgroups.

$\cdot$	e	r_1	r_2	v	u	d
e	e	r_1	r_2	v	u	d
r_1	r_1	r_2	e	u	d	v
r_2	r_2	e	r_1	d	v	u
v	v	d	u	e	r_2	r_1
u	u	v	d	r_1	e	r_2
d	d	u	v	r_2	r_1	e

Let $H = \{e, r_1, r_2\}$. Despite D_3 being nonabelian, this subgroup is normal. For any $g \in H$, the coset $gH = Hg$. If $g \notin H$, $gH = \{u, v, d\} = Hg$. Therefore *H* is a normal subgroup of *G*. We stress that the group is nonabelian, and so element by element it is not generally true that $gh = hg$. What is true is that the left and right cosets have the same elements.

Now let $K = \{e, u\}$. For an element such as r_1 we have $r_1K = \{r_1, d\}$ whereas $Kr_1 = \{r_1, v\}$. Therefore *K* is not a normal subgroup of *G*.

□

Theorem 37. *If a subgroup H has index of 2 in the group G then H is a normal subgroup.*

Proof.
We suppose that $(G : H) = 2$. Thus there are two cosets of $H : H$ and every other element of *G*, which we will denote by H^C. If $g \in H$ we have $gH = H = Hg$. If $g \in H^C$ we have $gH = H^C = Hg$. By this we see that *H* must be a normal subgroup.

□

The above theorem shows that nonabelian subgroups can be normal. For this we only need to consider A_5 in S_5. These are both nonabelian groups, but A_n is of index two in S_n for all $n \geq 2$.

Theorem 38. The following are equivalent, and so may be used as definitions of a normal subgroup. For subgroup *H* in *G*

1. $gH = Hg$ for all $g \in G$.
2. $g^{-1}Hg = H$ for all $g \in G$.
3. For any $g \in G$ and $h \in H$ $g^{-1}hg \in H$.
4. For any $g \in G$ and every $h \in H$ there exists a $k \in H$ such that $g^{-1}hg = k$.

Proof. The proof is left as an exercise.

□

2.9.3 Exercises

1. Find the cosets of $\langle 4 \rangle$ in $_{28}$
2. Find the cosets of $\{e, u\}$ in D_4.
3. Let p and q be prime numbers. Prove that every proper subgroup of $_{pq}$ is cyclic.
4. Let $H = \{e, (1, 2), (3, 4), (1, 2)(3, 4)\}$. Determine if *H* is a normal subgroup in S_4.
5. Let *H* and *K* be normal subgroups of *G* and define the set
   $HK = \{hk \mid h \in H \text{ and } k \in K\}$. Prove that *HK* is a normal subgroup of *G*.
6. If $\phi : G \to G'$ is a group homomorphism, prove that $\ker \phi$ is a normal subgroup of *G*.

7. Let H be a the subgroup of S_4 generated by the cycle $(1, 2, 3, 4)$. Determine if H is a normal subgroup of S_4.
8. Prove Theorem 38.
9. For group homomorphism $\phi : G \to G'$ show that if H is a normal subgroup of G then $\phi(H)$ is a normal subgroup of $\phi(G)$.
10. For group homomorphism $\phi : G \to G'$ show that if K is a normal subgroup of $\phi(G)$ then $\phi^{-1}(K)$ is a normal subgroup of G.

2.10 Quotient Groups

We will now see some of the motivation for singling out the concept of a normal subgroup. In a certain sense, we can think of dividing a group by a normal subgroup (or factoring out a normal subgroup). To this end we will define a binary operation between cosets of H. In fact we will define $(xH) \cdot (yH) = (xy)H$. This binary operation makes sense due to H being a normal subgroup: $Hy = yH$. This allows us to start with $xHyH$ and rewrite it as $xyHH$. One issue with this definition is that a single coset can be represented in different ways. For example, if $H = \{0, 2, 4, 6\}$ in $\mathbb{Z}_8$, then this coset may be represented as $H = 2 + H = 4 + H = 6 + H$. The binary operation that we have defined on cosets must be able to account for these different representations of the same coset.

Theorem 39.

Let H be a normal subgroup of the group G. The binary operation between cosets defined by $(xH) \cdot (yH) = (xy)H$ is well defined.

Proof.
We begin by noting that for any $x, x' \in G$, $xH = x'H$ if and only if there is a $h \in H$ such that $x' = xh$. We consider the binary operation on the cosets xH and yH, and compare this to the binary operation different representations $x'H$ and $y'H$ of the same cosets, i.e. $xH = x'H$ and $yH = y'H$.

$$x'H \cdot y'H = xh_1H \cdot yh_2H = xh_1yh_2H$$

Since H is a normal subgroup there exists an $h_3 \in H$ such that $h_1y = yh_3$.

$$xh_1yh_2 = xyh_3h_2H = xyh_4H = xyH = xH \cdot yH$$

Therefore the binary operation is well defined.

□

Theorem 40.
For the group G, the set of left cosets of H is a group under the binary operation $(xH) \cdot (yH) = (xy)H$.

Proof.
We check that the axioms of a group are satisfied:

- Associativity is inherited from the group *G* and for all $x, y, z \in G$ we have $(xH)((yH)(zH)) = ((xH)(yH))(zH)$.
- The identity element is the coset containing e, namely *H* itself. For any $x \in H$ we have $H \cdot xH = eH \cdot xH = xH = xH \cdot eH$.
- By the multiplication structure $xHyH = xyH$ it is clear that the inverse of aH is the coset $a^{-1}H$.

□

Definition: The group in the previous theorem is called the *quotient group* of *G* by *H* and is written *G/H*. This is also called the *factor group* of *G* by *H*.

□

WARNING:

In order to construct a quotient group, we must use a normal subgroup. If we attempt to use any other subgroup, then while left and right cosets can be formed, the binary operation $xH \cdot yH = xyH$ will no longer be well defined.

□

One of the easiest ways to see the quotient group of a finite group of low order is to arrange the group *G* in terms of the cosets of *H*. By treating each coset as a block in the group table, we can see the form of the quotient group *G*/*H*.

Example:

We will examine the quotient group Z_{12}/H with the normal subgroup $H = \langle 4 \rangle$. We will begin by rearranging the group table of Z_{12} by the cosets of *H*:

$+$	0	4	8	1	5	9	2	6	10	3	7	11
0	0	4	8	1	5	9	2	6	10	3	7	11
4	4	8	0	5	9	1	6	10	2	7	11	3
8	8	0	4	9	1	5	10	2	6	11	3	7
1	1	5	9	10	2	6	11	3	7	4	8	0
5	5	9	1	6	10	2	7	11	3	8	0	4
9	9	1	5	2	6	10	3	7	11	0	4	8
2	2	6	10	3	7	11	4	8	0	5	9	1
6	6	10	2	7	11	3	8	0	4	9	1	5
10	10	2	6	11	3	7	0	4	8	1	5	9
3	3	7	11	4	8	0	5	9	1	6	10	2
7	7	11	3	8	0	4	9	1	5	10	2	6
11	11	3	7	0	4	8	1	5	9	2	6	10

We see that there are four cosets: $H, 1 + H, 2 + H,$ and $3 + H$. Further inspection of the group table, particularly the fact that $(1 + H)$ is a generator with $(1 + H) + (1 + H) = 2 + H$ and $(1 + H) + (1 + H) + (1 + H) = 3 + H$ reveals that the quotient group Z_{12}/H is isomorphic to $\mathbb{Z}_4$.

□

Example:

Another normal subgroup of $\mathbb{Z}_{12}$ is $K = \langle 2 \rangle$. The cosets are

$$\{0, 2, 4, 6, 8, 10\} \text{ and } \{1, 3, 5, 7, 9, 11\}.$$

We rearrange the group table into these cosets. The quotient group $\mathbb{Z}_{12}/K$ is isomorphic to $\mathbb{Z}_2$.

$\cdot$	0	2	4	6	8	10	1	3	5	7	9	11
0	0	2	4	6	8	10	1	3	5	7	9	11
2	2	4	6	8	10	0	3	5	7	9	11	1
4	4	6	8	10	0	2	5	7	9	11	1	3
6	6	8	10	0	2	4	7	9	11	1	3	5
8	8	10	0	2	4	6	9	11	1	3	5	7
10	10	0	2	4	6	8	11	1	3	5	7	9
1	1	3	5	7	9	11	2	4	6	8	10	0
3	3	5	7	9	11	1	4	6	8	10	0	2
5	5	7	9	11	1	3	6	8	10	0	2	4
7	7	9	11	1	3	5	8	10	0	2	4	6
9	9	11	1	3	5	7	10	0	2	4	6	8
11	11	1	3	5	7	9	0	2	4	6	8	10

□

Example:

We consider the dihedral group D_3 of symmetries of an equilateral triangle and have seen that the subgroup of rotations $H = \{e, r_1, r_2\}$ is a normal subgroup. The other coset is $uH = \{u, v, d\}$, which has alternative representations of vH and dH, and we see that the quotient group D_3/H has the following structure, isomorphic to $\mathbb{Z}_2$:

$\cdot$	H	uH
H	H	uH
uH	uH	H

□

Example:

Consider the group $\mathbb{Z}$ with normal subgroup $H = 3\mathbb{Z} = \{\cdots, -6, -3, 0, 3, 6, \cdots\}$. These are two infinite groups. We see that the quotient group $\mathbb{Z}/H$ is finite. The cosets of H are $H, 1+H, 2+H$. By the mapping $\phi : \mathbb{Z}_3 \to \mathbb{Z}/H$ defined by $\phi(x) = x + H$ we see that $\mathbb{Z}/H \cong_3$.

In general for any positive integer n $\mathbb{Z}/(n\mathbb{Z}) \cong \mathbb{Z}_n$.

□

2.10.1 Theorems Regarding Quotient Groups

The following theorems were illustrated by the examples above.

Theorem 41. *Let H be a normal subgroup of G. If G is finite then the order of the quotient group G/H is the order of G divided by the order of H : $|G/H| = |G|/|H|$.*

Proof.
We have seen that the normal subgroup H partitions *G* into $|G|/|H|$ equivalence classes. This is the total number of distinct cosets of *H*.

□

Theorem 42.

If H is a normal subgroup of G then the mapping $\phi : G \to G/H$ defined by $\phi(g) = gH$ is an onto homomorphism.

Proof.
It is clear that ϕ is an onto mapping as *G/H* consists of the left cosets of *H*. We now show that ϕ is a homomorphism. Let $x, y \in G$ and consider $\phi(xy) = (xy)H = xH \cdot yH = \phi(x) \cdot \phi(y)$. Therefore ϕ is a homomorphism.

□

Theorem 43.

If G is cyclic and H is any subgroup, then G/H is a cyclic group.

Proof.
Let a be the generator of the cyclic group G, i.e. $\langle a \rangle = G$. We claim that aH is a generator of G/H. We have seen that the mapping $\phi : G \to H$ defined by $\phi(g) = gH$ is an onto homomorphism. By examining $\phi(a^k) = a^k H$ we see that every element of G/H has the form $a^k H$ for some $k \in \mathbb{Z}$. By the binary operation we know $a^k H = (aH)^k$, so aH generates G/H.

□

Theorem 44. *Let H be a normal subgroup of G. The quotient group G/H is abelian if and only if H contains every element of the form $xyx^{-1}y^{-1}$ for $x, y \in G$.*

2.10.2 Simple Groups

Just as the prime numbers are building blocks of the positive integers, in that every

Definition:

A nontrivial group is *simple* if its only proper normal subgroup is $\{e\}$.

□

Theorem 45. *The alternating group A_n is simple for all $n \geq 5$.*

Proof. We present a sketch of the proof here, with individual steps left as exercises.

1. First show that every cycle of length 3 is an element of A_n for $n \geq 3$.
2. Next show that these cycles of length 3 generate A_n for $n \geq 3$
3. Show that every cycle of length 3 is generated by the particular cycles (a, b, i) for fixed a, b with $1 \leq a, b, i \leq n$ for $n \geq 3$.
4. For $n \geq 3$ show that if H is a normal subgroup of A_n that contains a cycle of length 3 then $H = A_n$.
5. Let *H* be a nontrivial normal subgroup of A_n for $n \geq 5$. Consider all of the possible forms of the elements in *H* and show why *H* must contain a cycle of length 3.

□

2.10.3 Exercises

1. Let $H = \{0, 3, 6, 9, 12\}$. Determine the quotient group $\mathbb{Z}_{15}/H$ and write the group table.
2. Prove that if *H* and *N* are normal subgroups of a group *N* with $N \subseteq H$ then H/N is a normal subgroup of G/N and the following isomorphism holds:
$$(G/N)/(H/N) \cong G/H$$
3. Prove that every cycle of length 3 is an element of A_n for $n \geq 3$.
4. Prove that cycles of length 3 generate A_n for $n \geq 3$
5. Prove that every cycle of length 3 is generated by the particular cycles (a, b, i) for fixed a, b with $1 \leq a, b, i \leq n$ for $n \geq 3$.

6. For $n \geq 3$ show that if H is a normal subgroup of A_n that contains a cycle of length 3 then $H = A_n$.
7. Use the preceeding three problems to show why A_n is simple for $n \geq 5$.

2.11 Direct Products

Back when you first studied arithmetic, you saw how multiplication is used to form large numbers from factors that could be quite small. The numbers 2 and 3 are not that large, but the product $2 \cdot 2 \cdot 3 \cdot 3 = 36$ is greater, and $2^{10} = 1024$ is greater still. In this section we will look at a process that allows us to build larger groups by "multiplying" smaller ones together. This process relies upon the Cartesian product.

Theorem 46. *Let G_1 and G_2 be groups. Define a binary operation on $G_1 \times G_2$ by $(a_1, a_2) \cdot (b_1, b_2) = (a_1 b_1, a_2 b_2)$. The set $G_1 \times G_2$ is a group under this binary operation.*

Proof. We begin by noting that the binary operation is defined in terms of each group that forms the product. Thus the a_i do not interact with the b_i. As a result the group $G_1 \times G_2$ will inherit its group structure from its components. As always we must check that the group axioms hold.

- The binary operation is associative because for every $(a_1, a_2), (b_1, b_2), (c_1, c_2) \in G_1 \times G_2$

$$((a_1, a_2) \cdot (b_1, b_2)) \cdot (c_1, c_2) = (a_1 b_1, a_2 b_2) \cdot (c_1, c_2) = ((a_1 b_1)c_1, (a_2 b_2)c_2)$$

Now by the associativity of the groups G_1 and G_2 we are allowed to say

$$((a_1b_1)c_1, (a_2b_2)c_2) = (a_1(b_1c_1), a_2(b_2c_2))$$

and it is clear that this element is equal to $(a_1, a_2) \cdot ((b_1, b_2) \cdot (c_1, c_2))$.

- Since G_1 is a group is has an identity element e. The group G_2 also has an identity, which we will denote E to distinguish it from e. For any $(a_1, a_2) \in G_1 \times G_2$ we have $(a_1, a_2) \cdot (e, E) = (a_1 \cdot e, a_2 \cdot E) = (a_1, a_2) = (e \cdot a_1, E \cdot a_2) = (e, E) \cdot (a_1, a_2)$. Thus (e, E) is the identity of $G_1 \times G_2$.
- For the element $(a_1, a_2) \in G_1 \times G_2$ we know that $a_1 \in G_1$ and $a_2 \in G_2$. These groups elements have inverses a_1^{-1} and a_2^{-1} and so $(a_1, a_2) \cdot (a_1^{-1}, a_2^{-1}) = (a_1a_1^{-1}, a_2a_2^{-1}) = (e, E) = (a_1^{-1}a_1, a_2^{-1}a_2)$.

□

Definition: The group defined in the previous theorem is the *direct product of G_1 and G_2*.

□

The direct product of G_1 and G_2 inherits much of its structure from G_1 and G_2. By a basic counting argument we can see that the order of $G_1 \times G_2$ is the product of the orders of G_1 and G_2. If both of the groups are abelian, then the direct product is abelian. We might ask, if G_1 and G_2 are both cyclic, then is the direct product $G_1 \times G_2$ also cyclic?

Example: Consider the group $\mathbb{Z}_2 \times \mathbb{Z}_3$. We will look at the element $(1, 1) \in \mathbb{Z}_2 \times \mathbb{Z}_3$. In the discussion that follows it is key to remember where each of these elements 1 are coming from. One of them is an element of $\mathbb{Z}_2$ and the other is an element of $\mathbb{Z}_3$. We will add this element $(1, 1)$ to itself.

$(1, 1)$

$(1, 1) + (1, 1) = (1 + 1, 1 + 1) = (0, 2)$ (since the $1 + 1$ in the first coordinate occurs in the group $\mathbb{Z}_2$).

$(1, 1) + (1, 1) + (1, 1) = (1, 1) + (0, 2) = (1, 0)$ (since the $1 + 2$ in the second coordinate occurs in the group $\mathbb{Z}_3$

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

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

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

This shows us a few things. The order of the element $(1,1)$ is 6. We also know that the order of the group $\mathbb{Z}_2 \times \mathbb{Z}_3$ is 6. Thus $\mathbb{Z}_2 \times \mathbb{Z}_3$ is a cyclic group of order 6 with generator $(1,1)$. For any natural number, up to isomorphism there is only one cyclic group of order n. By these considerations $\mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6$. For an explicit isomorphism, we could map the generator we found of $\mathbb{Z}_2 \times \mathbb{Z}_3$. to a generator of $\mathbb{Z}_6$.

□

In this particular case we see that the cyclic structure of $\mathbb{Z}_2$ and $\mathbb{Z}_3$ transferred to the direct product $\mathbb{Z}_2 \times \mathbb{Z}_3$, so it can happen. The question is if this is always the case. Another example will help to answer this.

Example:

Consider the direct product $G = \mathbb{Z}_2 \times \mathbb{Z}_2$. The elements of $G = \{(0,0),(1,1),(1,0),(0,1)\}$ and its group table is:

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

It is clear from the table (and we could have determined this without the table) that for all elements $(x,y) \in G$, $(x,y)+(x,y) = (0,0)$. The order of any element in the group is at most two, but the order of G is four. This shows us that *G* is not a cyclic group, and so we have answered our question that $\mathbb{Z}_n \times \mathbb{Z}_m$ is not necessarily a cyclic group.

It is worthwhile to note that we have previously seen the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ This is a group of order four in which all elements other than the identity have order two. The other group of order four that we have encountered is the Klein four group. The mapping $\phi : \mathbb{Z}_2 \times \mathbb{Z}_2 \to V$ given by $\phi((0,0)) = e, \phi((1,1)) = a, \phi((1,0)) = b, \phi((0,1)) = c$ is one-to-one and onto. A check of all of the possibilities shows that this mapping is an isomorphism, so $\mathbb{Z}_2 \times \mathbb{Z}_2 \cong V$.

These two examples gave us different results. In both cases we started with the direct product of two cyclic groups. In one case the direct product was cyclic, but in the other case the direct product was not cyclic. This demonstrates that the direct product is a useful way to construct new groups from old ones. But we would be right to ask how do we know before forming the direct product if we are going to end up with a cyclic group. Are there are any conditions to look for that cause $\mathbb{Z}_n \times \mathbb{Z}_m$ to be cyclic?

We could form several direct products and see if we noticed a pattern. The group $\mathbb{Z}_3 \times \mathbb{Z}_5$ is cyclic, but $\mathbb{Z}_3 \times \mathbb{Z}_9$ is not. The group $\mathbb{Z}_6 \times \mathbb{Z}_7$ is cyclic, but $\mathbb{Z}_6 \times \mathbb{Z}_{14}$ is not. If we formed enough direct products, we would realize the following.

Theorem 47. *The group $\mathbb{Z}_n \times \mathbb{Z}_m$ is isomorphic to $\mathbb{Z}_{mn}$ (and thus cyclic) if and only if m, n are relatively prime.*

Proof. Recall for this proof that two positive integers are relatively prime if their greatest common divisor is 1. We begin by considering $(1,1) \in \mathbb{Z}_m \times \mathbb{Z}_n$. We add this element to itself repeatedly and observe that if we add it to itself $n, 2n, 3n$ or any multiple of n times, then the result is an element of the form $(x, 0)$. In a similar fashion, if we add the element $(1,1)$ to itself $m, 2m, 3m$ or any multiple of m times, then the result is an element of the form $(0, y)$.

Any common multiple k of both m and n will result in $\overbrace{(1,1)+(1,1)+\cdots(1,1)}^{k} = (0,0)$. The smallest such k is the least common multiple of m, n, denoted $\text{lcm}(m,n)$.

Suppose that m, n are relatively prime. Then $\gcd(m,n) = 1$ and by the equation

$$mn = \text{lcm}(m,n)\gcd(m,n)$$

we see that $\text{lcm}(m,n) = mn$. This shows that the element $(1,1)$ has order mn in a group with order mn. So $(1,1)$ is a generator of $Z_m \times \mathbb{Z}_n$ and this is a cyclic group. As there is only one cyclic group of order mn, $Z_m \times \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_{mn}$.

For the converse statement, suppose that m, n are not relatively prime. Thus $\gcd(m,n) = d > 1$ and by the equation $mn = \text{lcm}(m,n)\gcd(m,n)$ we see that $mn/d = \text{lcm}(m,n)$. Any element $(r,s) \in \mathbb{Z}_m \times \mathbb{Z}_n$ has order at most mn/d, which is less than mn. There is no generator of $\mathbb{Z}_n \times \mathbb{Z}_m$ so it is not cyclic, and $\mathbb{Z}_m \times \mathbb{Z}_n \ncong \mathbb{Z}_{mn}$.

□

Example:

By theorem 47 we see that there are a number of ways to express isomorphic groups:

$$\mathbb{Z}_6 \times \mathbb{Z}_{10} \cong \mathbb{Z}_6 \times \mathbb{Z}_5 \times \mathbb{Z}_2 \cong \mathbb{Z}_{30} \times \mathbb{Z}_2 \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_3$$

□

2.11.1 Direct Product of Several Groups

We can extend the construction of a direct product to more than two groups. There is a little bit of notation that needs to be introduced, but the overall process is the same as the direct product of two groups. We form a Cartesian product, and then we define a binary operation on the product.

Definition: The *Cartesian product of sets* $S_1, S_2, \cdots, S_n$ is the set of all ordered pairs $(a_1, a_2, \cdots, a_n)$ where $a_i \in S_i$ for $i = 1, 2, \cdots, n$ and is denoted $S_1 \times S_2 \times \cdots \times S_n = \prod_{i=1}^{n} S_i$.

□

This definition is a generalization of our Cartesian product for two sets. We are now allowed to use several sets to form a product. This allows to form more than the ordered pair (a_1, a_2). We now have the ability to form an ordered n-tuple $(a_1, a_2, \cdots, a_n)$. The word "ordered" is key, as changing this will result in a different point of $S_1 \times S_2 \times \cdots \times S_n$. The generalized Cartesian product is now used to form a direct product of several groups.

Theorem 48. *Let* $G_1, G_2, \cdots, G_n$ *be groups. For* $(a_1, a_2, \cdots, a_n), (b_1, b_2, \cdots, b_n) \in \prod_{i=1}^{n} G_i$. *Define*

$$(a_1, a_2, \cdots, a_n) \cdot (b_1, b_2, \cdots, b_n) = (a_1 b_1, a_2 b_2, \cdots, a_n b_n).$$

Then $\prod G_i$ *is a group.*

Proof. This result makes sense, but we will be somewhat careful in proving this because the proof uses a strategy that we have not seen for awhile. Since we are proving a statement regarding the natural numbers, we will use mathematical induction for the proof.

We have already proved that $G_1 \times G_2$ forms a group under the defined binary operation. Thus our induction proof is anchored. Suppose that $G_1 \times G_2 \times \cdots \times G_k$ is a group G. It is straightforward to see that $G_1 \times G_2 \times \cdots \times G_k \times G_{k+1} = G \times G_{k+1}$. Since this is a Cartesian product of two groups, it too is a direct product and a group.

□

Theorem 49. *Let $(a_1, a_2, \cdots, a_n) \in \prod G_i$. If a_i is of finite order s_i in each G_i, then the order of the element $(a_1, a_2, \cdots, a_n)$ is equal to the least common multiple of all the s_i.*

Proof. We induct on the number of groups G_i in the direct product. We have already seen that this theorem is true for $G_1 \times G_2$. By induction we suppose that the order of $(a_1, a_2, \cdots, a_k)$ is $\text{lcm}(s_1, s_2, \cdots, s_k)$ in $G_1 \times G_2 \times \cdots \times G_k$. Now we consider the order of $(a_1, a_2, \cdots, a_k, a_{k+1})$ in $G_1 \times G_2 \times \cdots \times G_k \times G_{k+1}$. Let $s = \text{lcm}(s_1, s_2, \cdots, s_k)$ and denote $G_1 \times G_2 \times \cdots G_k = G$. Since $G_1 \times G_2 \times \cdots \times G_k \times G_{k+1} \cong G \times G_{k+1}$, $(a_1, a_2, \cdots, a_k)$ has order s in G and a_{k+1} has order s_{k+1}, the element $(a_1, a_2, \cdots, a_k, a_{k+1})$ has order $\text{lcm}(s, s_{k+1}) = \text{lcm}(s_1, s_2, \cdots, s_k, s_{k+1})$.

□

Example:

Find the order of $(2, 6, 5) \in \mathbb{Z}_{12} \times \mathbb{Z}_{30} \times \mathbb{Z}_{20}$

First we find the order of each element in its respective group. 2 is of order six in $\mathbb{Z}_{12}$ 6 is of order five in $\mathbb{Z}_{30}$ and 5 is of order four in $\mathbb{Z}_{20}$. The order of $(2, 6, 5)$ is the $\text{lcm}(6, 5, 4) = 60$.

□

2.11.2 Finitely Generated Abelian Groups

We now move on to see what else there is to do with the direct product construction. We will be able to use this to classify a certain kind of abelian group.

Definition: A group is *finitely generated* if the group can be presented in terms of a finite list of generators and relations.

□

Most of the groups that we have run into are finitely generated. Cyclic groups are generated by a single element of a group. Dihedral groups can be expressed in terms of two generators, which are related by three relations. Groups such as $\mathbb{Q}$ or $\mathbb{R}$ under addition is not finitely generated. If we turn our focus to finitely generated abelian groups, all of these types of groups can be described in terms of direct products.

Theorem 50. *Fundamental Theorem of Finitely Generated Abelian Groups Every finitely generated abelian group is isomorphic to a finite direct product of cyclic groups.*

The proof of this theorem is beyond the level of this book. Since we have a classification of cyclic groups, a relatively straightforward corollary gives us a more explicit description of finitely generated abelian groups.

Corollary 51. *Every finitely generated abelian group is a direct product of a finite number of $\mathbb{Z}$ and a finite number of finite cyclic groups of the form $\mathbb{Z}_{p_i}^{s_i}$ where p_i is prime number and s_i is a positive integer.*

Proof. Let G be a finitely generated abelian group. We have seen that every cyclic group is isomorphic to $\mathbb{Z}$ or to $\mathbb{Z}_n$. The Fundamental Theorem of Finitely Generated Abelian Groups tells us that

$$G \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \mathbb{Z}_{n_k} \times \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}.$$

We use the prime factorization of each $n = p_1^{s_1} p_2^{s_2} \cdots p_j^{s_j}$. Since $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$ when $\gcd(m, n) = 1$, we can express $\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{s_1}} \times \mathbb{Z}_{p_2^{s_2}} \cdots \mathbb{Z}_{p_j^{s_j}}$. We rewrite each of the $\mathbb{Z}_{n_i}$ in this way.

□

Note:

The prime numbers p_i for the cyclic groups $\mathbb{Z}_{p_i^{s_i}}$ may be repeated. The direct product above is unique up to rearrangement of the groups $\mathbb{Z}_{p_i^{s_i}}$ and $\mathbb{Z}$. The usefulness of corollary 51 is that is can be used to determine all abelian groups of a particular order. The theorem takes a question regarding groups and turns it into a question involving the prime factorization of a number.

□

Example: How many abelian groups are there of order 8?

We factor $8 = 2^3$. We now partion this factorization in every way possible. By the fundamental theorem of finitely generated abelian groups, the following are the finite groups of order 8:

$$Z_8 \qquad \mathbb{Z}_4 \times \mathbb{Z}_2 \qquad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$$

□

Example: How many abelian groups are there of order 180?

We see that $180 = 2^2 3^2 5$. Each of the squared primes may be expressed in two different ways, so there are a total of four abelian groups of order 180.

$\mathbb{Z}_4 \times \mathbb{Z}_9 \times \mathbb{Z}_5$

$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_9 \times \mathbb{Z}_5$

$\mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5$

$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5$

□

2.11.3 Exercises

1. Prove that the direct product of G_1 and G_2 is abelian if and only if G_1 and G_2 are both abelian groups.
2. Prove that for groups G_1 and G_2 the following isomorphism holds: $G_1 \times G_2 \cong G_2 \times G_1$.
3. A positive integer m is said to be "square free" if m is not divisible by the square of any prime. Prove that every abelian group of order m is cyclic.

4. If $n = p_1^{k_1} p_2^{k_2}$ where p_1, p_2 are distinct prime numbers, how many abelian groups of order n exist?
5. List the elements of $\mathbb{Z}_3 \times \mathbb{Z}_6$. Is this a cyclic group?
6. Find the order of $(3, 4, 5, 6)$ in $\mathbb{Z}_{12} \times \mathbb{Z}_{20} \times \mathbb{Z}_{30} \times \mathbb{Z}_{24}$.
7. Explain whether the groups $\mathbb{Z}_{20} \times \mathbb{Z}_6$ and $\mathbb{Z}_1 2 \times \mathbb{Z}_{10}$ are isomorphic.
8. List all abelian groups of order 540.
9. The *torsion subgroup* of the group *G* is the set of all elements of *G* of finite order. Prove that this is indeed a subgroup.
10. Find the order of the quotient group $\mathbb{Z}_{12} \times \mathbb{Z}_{20} / \langle (1, 1) \rangle$.
11. What is the order of the element $(3, 3) + \langle (1, 1) \rangle$ in the quotient group $\mathbb{Z}_4 \times \mathbb{Z}_4 / \langle (1, 1) \rangle$.

2.12 Catalog of Finite Groups

Throughout this chapter we have encountered several different groups. Before moving on to other topics in abstract algebra, it would be worthwhile to collect ourselves and summarize what we have learned about finite groups of low order. We will examine up to isomorphism nearly all groups of order less than 20. In general it is a difficult to tell how many groups there are of a given order, and only partial result are known for some values of n . For what follows recall that:

- S_n is the symmetric group on n elements of order $n!$,
- A_n is the alternating group of order $n!/2$,
- D_n is the dihedral group of order $2n$, and
- $\mathbb{Z}_n$ is the cyclic group of order n .

1. For $n = 1$ there is only the trivial group $\{e\}$. This is considered a cyclic group as well as A_2
2. For $n = 2$ there is only the cyclic group $\mathbb{Z}_2 \cong S_2 \cong D_1$.
3. For $n = 3$ there is only the cyclic group $\mathbb{Z}_3 \cong A_3$.
4. For $n = 4$ there are two groups, both of which are abelian:
   a) The cyclic group $\mathbb{Z}_4$
   b) The Klein four-group $V \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \cong D_2$.
5. For $n = 5$ there is only the cyclic group $\mathbb{Z}_5$.
6. For $n = 6$ there are two groups.
   a) $\mathbb{Z}_6$ is cyclic and thus abelian
   b) $S_3 \cong D_3$ is nonabelian.

   For any even number n greater than or equal to 6 we will have both $\mathbb{Z}_n$ and $D_{n/2}$ as nonisomorphic groups.
7. For $n = 7$ there is only the cyclic group $\mathbb{Z}_7$.
8. For $n = 8$ there are five nonisomorphic groups

a) Abelian groups of order 8:
   i. $\mathbb{Z}_8$ is a cyclic group
   ii. $\mathbb{Z}_2 \times \mathbb{Z}_4$
   iii. $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$

b) Nonabelian groups of order 8:
   i. D_4
   iv. Q is the group of quaternions. This group has generators $(-1), i, j, k$ where (-1) commutes with every element of the group, and the elements are related by $(-1)^2 = e$, $i^2 = j^2 = k^2 = ijk = -1$. This is nonabelian because $ij = k$, whereas $ji = (-1)k$.

9. For $n = 9$ there are two abelian groups:
   a) The cyclic group $\mathbb{Z}_9$
   b) The direct product $\mathbb{Z}_3 \times \mathbb{Z}_3$

10. For $n = 10$ there are two groups.
   a) $\mathbb{Z}_{10}$ is cyclic and thus abelian
   b) D_5 is nonabelian.

11. For $n = 11$ there is only the cyclic group $\mathbb{Z}_{11}$.

12. For $n = 12$ there are five nonisomorphic groups
    a) Abelian groups of order 12:
        i. $\mathbb{Z}_{12}$ is a cyclic group
        ii. $\mathbb{Z}_2 \times Z_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$
    b) Nonabelian groups of order 12:
        i. D_6 This has an element of order 6, which distinguishes it from the other nonabelian groups of order 12.
        i. A_4
        ii. The dicyclic group of order 12 is $\{a, b, c \mid a^3 = b^2 = c^2 = abc\}$

13. For $n = 13$ there is only the cyclic group $\mathbb{Z}_{13}$.
14. For $n = 14$ there are two groups.
    a) $\mathbb{Z}_{14}$ is cyclic and thus abelian
    b) D_7 is nonabelian.

15. For $n = 15$ there is only the cyclic group $\mathbb{Z}_{15}$.
16. For $n = 16$ there are 14 nonisomorphic groups.
    a) Abelian groups of order 16:
        i. $\mathbb{Z}_{16}$ is a cyclic group
        ii. $\mathbb{Z}_2 \times \mathbb{Z}_8$
        iii. $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4$
        iv. $\mathbb{Z}_4 \times \mathbb{Z}_4$
        v. $\mathbb{Z}_2 \times Z_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$

    b) Nonabelian groups of order 16:
        i. D_8
        ii. $\mathbb{Z}_2 \times Q$
        iii. $\mathbb{Z}_2 \times D_4$
        iv. And six other groups with descriptions beyond the scope of this book.

17. For $_2 \times Q$ there is only the cyclic group $\mathbb{Z}_{17}$.
18. For $^{n = 18}$ there are five nonisomorphic groups.
    a) Abelian groups of order 18:
        i. $\mathbb{Z}_{18}$ is a cyclic group
        ii. $\mathbb{Z}_3 \times \mathbb{Z}_6$
    b) Nonabelian groups of order 18:
        i. D_9
        ii. $S_3 \times \mathbb{Z}_3$
        iii. Generalized dihedral group of order 18

19. For $n = 19$ there is only the cyclic group $\mathbb{Z}_{19}$.

2.12.1 Exercises

1. Up to isomorphism, how many groups are there of order p where p is a prime?
2. Up to isomorphism, how many groups are there of order pq where p and q are distinct primes?
3. Up to isomorphism, how many groups are there of order pq^2 where p and q are distinct primes?
4. Up to isomorphism, how many groups are there of order $p^m q^n$ where p and q are distinct primes?
5. Though both are nonabelian groups of order 8, prove that D_4 is not isomorphic to Q.
6. Prove that A_4 is not isomorphic to the dicyclic group of order 12.

3 Field Theory

3.1 Introduction to Fields

We will now move on from our study of groups and expand our horizons to another algebraic structure. The rational numbers $\mathbb{Q}$ are equipped with two operations of addition and multiplication. We have already seen that the entire set of rational numbers with addition supports an abelian group structure, in which addition is defined as

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$$

where a, b, c, d are all integers with b, d nonzero. Of course we know from our study of arithmetic that addition is not the only thing we can do with fractions. There is also a way to multiply fractions in which we define

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

The nonzero rational numbers with this multiplication supports a second abelian group structure. Further consideration reveals that we also need to think about how our two operations of addition and multiplication interact with each other. In this case we have the distributive property $r(s + t) = rs + rt$.

This phenomenon is observed when dealing with the set of real numbers as well. The set $\mathbb{R}$ with addition is an abelian group. The set of nonzero real numbers $\mathbb{R}^*$ under multiplication forms an abelian group. The multiplication distributes across the addition for the real numbers.

What we have observed in both of these situations is a new algebraic structure. This structure builds upon our definition of group in a way that is helpful for further applications. In doing this we are able to abstract more areas of mathematics, now that we are allowing ourselves a set with two binary operations.

Definition:

A *field* is a set *F* with two binary operations $+, \cdot$

1. F with $+$ is an abelian group, with identity that we write 0.
2. F^*, the nonzero elements of *F*, with $\cdot$ is an abelian group, with identity that we write 1.
3. For all $r, s, t \in F$ we have the distributive property $r \cdot (s + t)$

□

Notation: Even though we use $+$ and $\cdot$ above, just as when we studied groups the "addition" and "multiplication" performed may not be our standard operations. The additive inverse of $a \in F$ will be written by $-a$. The multiplicative inverse of $a \in F^*$ will be written as $a^{-1} = \frac{1}{a}$. When the multiplication is clear from the context, we may not explicitly write $a \cdot b$, but instead ab. We use these symbols and conventions for convenience and connection to fields close to our intuition, such as $\mathbb{Q}$ and $\mathbb{R}$

□

We have already seen two important examples of fields. We will see two more that may be slightly more unfamiliar.

Example:

The set of complex numbers $\mathbb{C}$ forms a field under the operations of $(a + bi) + (c + di) = (a + b) + (c + d)i$ and $(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i$
Addition is commutative and we see that $0 + 0i$ is the additive identity with $-a - bi$ the additive inverse of $a + bi$.

Multiplication is commutative and we see that $1 + 0i$ is the multiplicative identity. For all nonzero complex numbers $a + bi$ we have multiplicative inverse $\frac{a-bi}{a^2+b^2}$

The demonstration that the distributive property holds is an exercise.

□

Example:

$\mathbb{Z}_p$ is a field when p is a prime number. The set of equivalence classes modulo p forms an abelian group $\mathbb{Z}_p$ under addition. The nonzero elements of $\mathbb{Z}_p$ form an abelian group under multiplication. The integers $1, 2, 3, \cdots, p-1$ are all relatively prime to p. Thus for any $1 \leq x \leq p$ there exist integers $k, n \in \mathbb{Z}$ such that $kx + np = 1$. Working modulo p we see that $(k \bmod p)(x \bmod p) = 1$. Thus $(k \bmod p)$ is the multiplicative inverse of x.

Since the integers possess the distributive property, this is inherited by $\mathbb{Z}_p$.

□

We will eventually see that there is a way to construct a field with order of p^k where p is any prime and k is any positive integer.

Example:

The following addition and multiplication tables display a field of order $2^2 = 4$.

$+$	0	1	α	$1+\alpha$
0	0	1	α	$1+\alpha$
1	1	0	$1+\alpha$	α
α	α	$1+\alpha$	0	1
$1+\alpha$	$1+\alpha$	α	1	0

$\cdot$	0	1	α	$1+\alpha$
0	0	0	0	0
1	0	1	α	$1+\alpha$
α	0	α	$1+\alpha$	1
$1+\alpha$	0	$1+\alpha$	1	α

□

Definition:

A *field homomorphism* is a mapping $\phi : F \to E$ where for all $x, y \in F$

$$\phi(x+y) = \phi(x) + \phi(y) \quad \text{and} \quad \phi(x \cdot y) = \phi(x) \cdot \phi(y).$$

A *field isomorphism* is a field homomorphism that is also one-to-one and onto.

□

3.1.1 Theorems Regarding Fields

One aspect of studying algebra in an axiomatic way is that many results that are "obvious" need to be proved from the given statements. We will see a few examples of these kinds of theorems here. For instance for any field the product of the additive identity with any field element gives us the additive identity. This gives us the familiar formula $0 \cdot a = 0$. Rather than saying that this statement is true because it is true, we will say it is true because we can prove it is true from the axioms for a field.

Theorem 52. *For the field F and any element $a \in F$, $0 \cdot a = a \cdot 0 = 0$.*

Proof. We begin with $0 \cdot a + 0 \cdot a$ and use the distributive property:

$$0 \cdot a + 0 \cdot a = (0 + 0) \cdot a = 0 \cdot a$$

Thus $0 \cdot a + 0 \cdot a - 0 \cdot a = 0 \cdot a - 0 \cdot a$ and $0 \cdot a = 0$. By the commutative property $0 \cdot a = a \cdot 0$.

□

Theorem 53. *For the field F and any element $a \in F$, $(-1)a = -a$.*

Proof. We begin with $a + (-1) \cdot a$ and use the distributive property:

$$a + (-1)a = (1 + -1) \cdot a = 0 \cdot a$$

and by the previous theorem $0 \cdot a = 0$. Since $a + (-1)a = 0$ we have $(-1)a = -a$.

□

Theorem 54. *For the field F and any elements $a, b \in F$, $(-a)(-b) = ab$.*

Proof. By the previous theorem we have $(-a)(-b) = (-1)a(-1)b$. We use the commutative property of multiplication and see $(-1)a(-1)b = (-1)(-1)ab = ab$.

□

Theorem 55. *Every field homomorphism is one-to-one or trivial.*

Proof. Let $\phi : F \to E$ be a field homomorphism. If $\ker\phi = 0$ then ϕ is one-to-one. Suppose that ϕ is not one-to-one. So there is an nonzero element $x \in \ker\phi$.

$$0 = 0 \cdot \phi(x^{-1}) = \phi(x)\phi(x^{-1}) = \phi(1)$$

For any $y \in F$ we have $\phi(y) = \phi(y \cdot 1) = \phi(y) \cdot \phi(1) = \phi(y) \cdot 0 = 0$ Since $F = \ker\phi$, the mapping ϕ is trivial.

□

Corollary 56. *Any onto field homomorphism is a field isomorphism.*

Proof. If we know that a field homomorphism ϕ is onto, then it is not trivial. By the previous theorem ϕ is one-to-one. Therefore ϕ is an isomorphism.

□

3.1.2 Exercises

1. Prove that the distributive property holds for complex numbers $\mathbb{C}$.
2. For the field F and any elementa $a, b \in F$, prove that if $a \cdot b = 0$ then $a = 0$ or $b = 0$.
3. Show that $(a + b)^p = a^p + b^p$ for p a prime and $a, b \in \mathbb{Z}_p$.
4. Prove that the fields $\mathbb{R}$ and $\mathbb{C}$ are not isomorphic.

3.2 Polynomials

Polynomials are found throughout basic mathematics. Finding the zeros of polynomials was one of the driving forces that led to the development of abstract algebra. We can intuitively think of a polynomial as an expression of the form $c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$, where x is an indeterminate. This indeterminate is to be thought of as an algebraic quantity, not a variable that takes on a value. This is a subtle but important distinction from the algebra that we may have encountered in a high school mathematics course. Rather than finding the solution to the equation $x^2 + 3x + 1 = 0$ we will be finding the zeros of the polynomial $f(x) = x^2 + 3x + 1$. We will find that the zeros of a polynomial are intimately linked to the field over which we are working. As a straightforward example of this, what are the zeros of the polynomial $x^2 - 5$? If only allow ourselves the possibility of rational zeros, then there are none. If we instead work over the field of real numbers, then $\pm\sqrt{5}$ are zeros of the polynomial. For mathematical precision we need to more carefully define a polynomial than what we have done above.

Definition:

A *polynomial* $f(x)$ *with coefficients in the field* F is an infinite formal sum $\sum_{i=0}^{\infty} c_ix^i = c_0 + c_1x + c_2x^2 + \cdots c_nx^n + \cdots$ where $c_i \in F$ and all but a finite number of $c_i = 0$.

□

Definition: The field elements c_i are called the *coefficients* of the polynomial $f(x)$.

The element x is called an *indeterminant* of the polynomial $f(x)$.

The largest value of i for which $c_i \neq 0$ is called the *degree* of the polynomial $f(x)$. We write this as $\deg f$.

□

Notation:

The set of all polynomials over the field *F* with indeterminant x is denoted by $F[x]$. This set has some additional structure, which we will explore in more depth in the final unit of the book.

□

Note:

A polynomial of the form $f(x) = c_0$ is a constant polynomial. Since c_0 is an element of the field F we can consider $F \subseteq F[x]$. Constant polynomials are all considered to have degree of zero.

□

We define polynomial addition and multiplication to match that from our other exposures to algebra. Addition is relatively straightforward to define formally. If $f(x) = \sum_{i=0}^{\infty} c_i x^i$ and $g(x) = \sum_{i=0}^{\infty} d_i x^i$ then $(f+g)(x) = \sum_{i=0}^{\infty} (c_i + d_i) x^i$. Multiplication is slightly more complicated to define, as we must account for all of the possible ways to obtain coefficients of a given power of x. $(fg)(x) = \sum_{i=0}^{\infty} e_i x^i$ where $e_i = \sum_{j=0}^{i} a_j b_{i-j}$. Here we are formally saying that the term x^3 is obtained in any of the following ways:

1. $1 \cdot x^3$
2. $x \cdot x^2$
3. $x^2 \cdot x$
4. $x^3 \cdot 1$

Things can get interesting when we work with fields other than $\mathbb{Q}$ or $\mathbb{R}$.

Example:

Calculate the product $(x+1)(x^2+x+1)$ over the field $\mathbb{Z}_2$.

The main thing to remember is that in the field $\mathbb{Z}_2$ the only coefficients are 0 and 1. Thus $(x+1)(x^2+x+1) = x \cdot (x^2+x+1) + 1 \cdot (x^2+x+1) = x^3 + x^2 + x + x^2 + x + 1 = x^3 + 1$

□

Theorem 57. *For nonzero polynomials* f, g *over a field* F

1. $\deg(f+g) \leq \max\{\deg f, \deg g\}$
2. $\deg(fg) = \deg f + \deg g$

Proof. Left as an exercise.

□

We can divide polynomials using a process that is similar the long division we use for integers. One reason for doing this is to determine the factors of a polynomial $f(x)$. In other words we want to find polynomials such that $f(x) = p(x)q(x)$.

Theorem 58 (Division Algorithm). *Let* f, g *be polynomials over the field* F *with* $g \neq 0$. *Then there exist unique polynomials* $q, r \in F[x]$ *such that* $f = qg + r$ *and* $r(x) = 0$ *or* $\deg r < \deg g$.

Proof. Consider the set of polynomials over F of the form $f - tg$ for some $t \in F[x]$, i.e. $R = \{f(x) - t(x)g(x) \mid t(x) \in F[x]\}$. There are two cases to consider: $0 \in R$ and $0 \notin R$.

If $0 \in R$ then there is a $t(x)$ such that $0 = f(x) - t(x)g(x) \Rightarrow f(x) = t(x)g(x)$. We set $q(x) = t(x)$ and $r(x) = 0$.

If $0 \notin R$, then we choose a polynomial in R of minimal degree and call this polynomial $r(x)$, we call $t(x) = q(x)$ and have: $f(x) = g(x)q(x) + r(x)$. If $m = \deg r \geq \deg g = n$ then there is an element $a \in F$ such that

$$r(x) - cx^{m-n}g(x) = (f(x) - g(x)q(x)) - cx^{m-n}g(x) = f(x) - g(x)(q(x) - cx^{m-n}) \in R$$

has degree less than or equal to $m - 1$. This contradicts the fact that $r(x)$ is of minimal degree in R. Therefore $\deg r < \deg g$.

We now demonstrate the uniqueness of the polynomials q, r. Suppose by way of contradiction that these polynomials are not unique. There are polynomials q, q, r, r such that $qg + r = f = qg + r$. We rearrange this equation and see $(q - q)g = r - r$. If $q - q \neq 0$ then by looking at the degrees of both sides of the equation we see

$$\deg(q - q) + \deg g = \max\{r, r\}.$$

Thus we have $\deg g \leq r$ or $\deg g \leq r$. But neither of these statements is true. Therefore the polynomials q, r are unique.

□

Again, we stress that our results are very much dependent upon the field that we're working over.

Example:

Perform the calculation $x^3 + 2x^2 + x + 1 \div (x + 2)$ over $\mathbb{Z}_3$.

We simply do some long division with our polynomials:

$$\begin{array}{rl} & x^2 + 1 \\ (x+2) & \overline{|x^3 + 2x^2 + x + 1} \\ - & \underline{(x^3 + 2x)} \\ & x + 1 \\ & \underline{-x + 2} \\ & 1 \end{array}$$

The quotient is $x^2 + 1$ with a remainder of $-1 = 2$

□

Theorem 59. *Let f be a polynomial in $F[x]$ and $a \in F$, then there is a unique polynomial $q(x) \in F[x]$ such that $f(x) = (x - a)q(x) + f(a)$*

Proof. We use the division algorithm with $g(x) = x - a$. This gives us $f(x) = (x - a)q(x) + r(x)$ where $\deg r < \deg(x - a)$. So $r(x)$ is a constant in *F*. Evaluation of $f(x)$ at a shows that this constant is $f(a)$.

□

Definition:

An element $a \in F$ is a *zero* of the polynomial $f(x) \in F[x]$ if $f(a) = 0$.

□

Corollary 60. *If $a \in F$ is a zero of the polynomial $f(x) \in F[x]$ then $x - a$ is a factor of $f(x)$.*

Proof. We use the previous theorem and see that if $f(a) = 0$ then $f(x) = (x - a)q(x) + 0$.

□

Of course the converse of this theorem is also true. If $x - a$ is a factor of $f(x)$ then a is a zero of $f(x)$.

Theorem 61. *A polynomial $f(x) \in F[x]$ of degree n has at most n zeros in F.*

Proof. The proof is by induction on the degree of the polynomial f. If f has degree of zero then the nonzero constant $c \in F$ has no zeros. Now suppose by induction that a polynomial of degree k has at most k zeros. Now let $f(x) \in F[x]$ be a polynomial of degree $k + 1$. There are two possibilities:

- The polynomial $f(x)$ has no zeros in *F*, in which case we are done.
- The polynomial $f(x)$ has a zero in *F*, in which case $x - a$ is a factor. We have $f(x) = (x - a)g(x)$ with $g(x)$ of degree k. By induction $g(x)$ has at most k zeros in *F*. Therefore $f(x)$ has at most $k + 1$ zeros in *F*.

□

There are at most n zeros in the field *F* for a polynomial of degree n over *F*. This is an upper bound, but is not always achieved for a given field. One field that is special in this regard is the field of complex numbers. For any polynomial $f \in [x]$ of degree n, f has exactly n zeros in $\mathbb{C}$. This fact is really just a statement of the Fundamental Theorem of Algebra. A final theorem in this section is also a corollary of our factor theorem. We (somewhat surprisingly) apply a theorem about factoring to obtain a result concerning the multiplicative structure of a finite field.

Theorem 62. *For a finite field F the group F^* under multiplication is a cyclic group.*

Proof. By the Fundamental Theorem of Finitely Generated Abelian Groups, $F^* \cong \mathbb{Z}_{p_1}^{n_1} \times \mathbb{Z}_{p_2}^{n_2} \times \cdots \times \mathbb{Z}_{p_k}^{n_k}$ where the p_i are prime numbers. Let $d = \text{lcm}({p_1}^{n_1}, {p_2}^{n_2}, \cdots, {p_k}^{n_k})$. For any element $a_i \in_{p_1}^{n_i}$ we have $a_i^d = 1$. Therefore every element $a \in F^*$ is a zero of $f(x) = x^d - 1$. There are at most d zeros in the field F, and there are ${p_1}^{n_1}{p_2}^{n_2} \cdots {p_k}^{n_k})$ elements in F. Therefore $d = {p_1}^{n_1}{p_2}^{n_2} \cdots {p_k}^{n_k})$ and the primes are relatively prime. Therefore $F^* \cong_d$.

□

3.2.1 Exercises

1. Prove that for nonzero polynomial f, g that
   a) $\deg(f+g) \leq \max\{\deg f, \deg g\}$
   b) $\deg(fg) = \deg f + \deg g$
2. Find the sum and product of $x^3 + 3x^2 + 2x + 4$ and $4x^3 + 3x^2 + x + 4$ over $\mathbb{Z}_5$.
3. Including 0 list the polynomials of degree 3 or less in $\mathbb{Z}_3[x]$.
4. Find all the zeros of $x^6 + 3x^4 + x^2 + 2x$ over $\mathbb{Z}_7$ and factor the polynomial.
5. Divide $x^4 - 2x^3 + 3x + 5$ by $x + 1$ over $\mathbb{Z}_7$.

3.3 Irreduciblity

From our early days in a high school algebra class, we should remember that some polynomials simply do not factor. For instance, the polynomial $x^2 + 9$ does not have any zeros in the field $\mathbb{R}$, and there is no way to write $x^2 + 9 = (x - a)(x - b)$ where $a, b \in \mathbb{R}$. We formalize this concept for polynomials over any field.

3.3.1 Basic Irreducibility Facts

Definition: Let $f(x) \in F[x]$ be a nonconstant polynomial. We say that $f(x)$ is *irreducible over F* if there are no polynomials $g(x), q(x) \in F[x]$ such that $f(x) = g(x)q(x)$ with degree of g, q less than that of f.

If f is not irreducible over *F* we say that it is *reducible over F*.

□

Irreduciblity is very much dependent upon the field that we are working over. The polynomial $x^2 - 3$ is irreducible over $\mathbb{Q}$, however it is reducible $x^2 - 3 = (x - \sqrt{3})(x + \sqrt{3})$ over $\mathbb{R}$.

□

If $f(x)$ has a zero in the field F then f is reducible. The converse of this statement is not true, however. Consider the polynomial $x^4 + 2x^2 + 1 = (x^2 + 1)^2$. This is clearly reducible, however there are no zeros in $\mathbb{R}$. When we consider polynomials of degree 2 or 3, then reducibility implies that there is a zero in the field *F*.

Example:
Is $f(x) = x^3 + 2x^2 + x + 1$ irreducible over $\mathbb{Z}_3$?

Since the degree of the polynomial is 3, if f is reducible, then it will factor into a polynomial of degree 1 and degree 2, or into three polynomials each of degree 1. In any case, if f is reducible of degree 3 then there will be a factor of the form $(x - a)$ with $a \in F$.

We check to see if this is the case by seeing if any of the field elements are zero: $f(0) = 1, f(1) = 2, f(2) = 2$. Since there is $a \in F$ such that $f(a) = 0$, the polynomial f is irreducible.

□

Definition:

A *monic* polynomial of degree n is a polynomial where x^n has coefficient of 1.

□

Example:

Find the monic irreducible polynomials over $\mathbb{Z}_3$ of degree 1 or 2.

Over the field $\mathbb{Z}_3$ we see that there are a total of 3 monic polynomials of degree 1: $x, x-1, x-2$ are all irreducible.

Now for degree 2 polynomials, there are a total of nine monic polynomials of degree 2 over $\mathbb{Z}_3$. It is clear that the constant term must be nonzero, for the three polynomials of the form x^2+ax are clearly reducible: $x^2+ax = x(x+a)$. Other reducible polynomials are $x^2+2 = (x+1)(x+2), (x+1)(x+1) = x^2+2x+1, (x+2)(x+2) = x^2+x+1$. So there are three irreducible polynomials of degree 3: x^2+1, x^2+2x+2, x^2+x+2. We can verify these are irreducible by seeing that none of them have zeros.

□

Theorem 63 (Eisenstein Irreducibility). *Let $f(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$ be a polynomial over the field of rational numbers $\mathbb{Q}$ with coefficients in the integers $\mathbb{Z}$. If there exists a prime number p such that:*

1. *p does not divide c_n,*
2. *p divides c_i for $0 \leq i \leq n-1$,*
3. *and p^2 does not divide c_0,*

then f is irreducible over $\mathbb{Q}$.

Proof. Suppose by way of contradiction that f is a polynomial that satisfies the criteria and is reducible. We write $f = gh$ with $g(x) = a_0 + a_1x + \cdots + a_nx^s$ and $h(x) = b_0 + b_1x + \cdots + b_nx^t$ and look at the coefficients c_i of f. Since $c_0 = a_0b_0$ we know that p divides a_0 or b_0 but not both since p^2 doesnï¿½t divide a_0b_0. We suppose that p divides a_0.

Now consider $c_1 = a_0b_1 + a_1b_0$. Since p divides $c_1, a_0,$ but not b_0 we know that p divides a_1. We continue this process and see that p divides a_i for all $1 \leq i \leq s$. Thus p divides c_n, which is a contradiction. Therefore f is irreducible over $\mathbb{Q}$.

□

Example:

Show that for a prime p the polynomial $\Phi_p(x) = x^{p-1} + \cdots + x^2 + x + 1$ is irreducible over $\mathbb{Q}$.

We note that $\Phi_p(x)(x-1) = x^p - 1$. If we replace x by $x+1$ we see that $\Phi_p(x+1)x = (x+1)^p - 1$. We write $\Phi_p(x+1) = \frac{1}{x}(x^p + px^{p-1} + \frac{p(p-1)}{2}x^{p-2} + \cdots + px = x^{p-1} + px^{p-2} + \frac{p(p-2)}{2}x^{p-3} + \cdots + p$. By the Eisenstein Irreducibility Criterion $\Phi_p(x+1)$ is irreducible over $\mathbb{Q}$. This implies that $\Phi_p(x)$ is irreducible over $\mathbb{Q}$. The polynomials $\Phi_p(x)$ are called *cyclotomic* ("circle splitting") polynomials due to their connection to the complex p th roots of unity.

□

3.3.2 Greatest Common Divisors

Definition: A *greatest common divisor* of two polynomials f, g of positive degree over a field F is any polynomial of maximum degree that divides both f and g.

□

We say "a" and not "the" greatest common divisor because if d divides f and g and c is a nonzero element of the field F, then cd is also a greatest common divisor. For instance, over $\mathbb{R}$ the polynomials $f(x) = x^2 - x - 6 = (x-3)(x+2)$ and $g(x) = x^2 - 6x + 9 = (x-3)^2$ have a greatest common divisor of $(x-3)$. However, they also have a greatest common divisor of $3 - x$. Other than this minor modification in our thinking, a greatest common divisor of polynomials over a field F works in much the same way that the greatest common divisor of two integers did. We see this especially in the next theorem.

Theorem 64. *If $f, g \in F[x]$ have positive degree and d is a greatest common divisor of f and g, then there exist polynomials $s, t \in F[x]$ such that $d = sf + tg$.*

Proof. Consider the set *A* of all polynomials of the form $sf + tg$ over the field *F*. Let $d' = s'f + t'g$ be of minimal degree. The polynomial d' divides every polynomial of the form $sf + tg$. If it did not divide h then $h = d'q + r$ with $\deg r < \deg d'$. However

$$r = h - d'q = (sf + tg) - (s'f + t'g)q = (s - s'q)f + (t - qt)g,$$

meaning that $r \in A$, a contradiction that $d' \in A$ is of minimal degree.

This shows that d' divides every element of *A*. Since $f = 1 \cdot f + 0 \cdot g$ and $g = 0 \cdot f + 1 \cdot g$ we know that d' divides f and g, and thereby d' divides d. We also see that because d divides f and g and $d' = s'f + t'g$, d divides d'. Therefore $d' = cd$ where $c \in F^*$. We obtain the desired equation by multiplying both sides of $d' = s'f + t'g$ by the inverse of c.

□

Definition:

For polynomials $f, g, q \in F[x]$ we say that $f \equiv g \bmod q$ if and only if q is a factor of $f - g$.

□

Theorem 65. *The previous definition is an equivalence relation.*

Proof. For $f, g, q \in F[x]$ we check the three conditions of an equivalence relation:

- $f \equiv f$ since q divides $f - f = 0$.
- If $f \equiv g$ then q divides $f - g$, then it follows that q divides $g - f$. Thus $g \equiv f$
- If $f \equiv g$ and $g \equiv h$ then q divides $f - g$ and q divides $g - h$. Thus q divides $(f - g) + (g - h) = f - h$. Therefore $f \equiv h$.

□

Notation:

We let $F[x]/(q)$ indicate the equivalence classes of $F[x]$ under $f \equiv \bmod q$. Let $\overline{f}$ denote the equivalence class of f.

□

Theorem 66. *Let $\overline{f}, \overline{g} \in F[x]/(q)$ as defined above, and binary operations $\overline{f} + \overline{g} = \overline{f + g}$, $\overline{f}\,\overline{g} = \overline{fg}$, the structure $F[x]/(q)$ is a field if and only if q is irreducible over F.*

Proof. If q is reducible then $q = hk$ where $h, k \in F[x]$ are of degree less than q. Thus $\overline{k} \neq \overline{0}$. However,

$$\overline{k} = \overline{h^{-1}}\overline{hk} = \overline{h^{-1}}\overline{0} = \overline{0}$$

so we have a contradiction.

Now suppose that q is irreducible over *F*. We need to check that the field axioms hold. The associativity and commutativity of the addition is inherited from the associativity and commutativity of addition in F. The class $\overline{q} = \overline{0}$ is the identity element. For any $\overline{f} \in F[x]/(q)$ we consider $\overline{q - f}$. Since $\overline{f} + \overline{q - f} = \overline{q} = \overline{0} \Rightarrow \overline{q - f} = -\overline{f}$. Thus $F[x]/(q)$ forms an abelian group under addition.

The associativity and commutativity of the multiplication is inherited from the associativity and commutativity of addition in $F[x]$. The element $\overline{1}$ is the multiplicative identity. Since q is irreducible, if f is not a multiple of q then the set of greatest common divisors of f and q contains 1. Thus by theorem 64 there exist $u, v \in F[x]$ such that $1 = uf + qv$. This shows that

$$\overline{1} = \overline{uf + qv} = \overline{uf} + \overline{0} = \overline{uv}$$

Thus any nonzero element $f \in F[x]/(q)$ has a multiplicative inverse.

It is a routine check and left as an exercise to see that the distributive property holds. Therefore $F[x]/(q)$ is a field.

□

3.3.3 Exercises

1. Show that $f(x) = x^2 + 2x + 2$ is irreducible over $\mathbb{Z}_3$
2. Is the polynomial $5x^8 - 6x^7 + 24x^3 + 18x^2 + 30x + 60$ irreducible over $\mathbb{Q}$?
3. Prove that the distributive property holds in $F[x]/(q)$ where $\overline{f} + \overline{g} = \overline{f + g}$, $\overline{f}\overline{g} = \overline{fg}$
4. Find all monic irreducible polynomials of degree 2 in $\mathbb{Z}_5[x]$.
5. Find the number of monic irreducible polynomials of degree 2 in $\mathbb{Z}_p[x]$, where p is a prime.
6. Prove that a polynomial f, irreducible over the field *F*, has a zero in *F* if and only if $\deg f = 1$.

3.4 Vector Spaces

In this section we borrow quite heavily from the realm of linear algebra. But rather than being interested in things such as matrices we are going to look at the abstract definition of a vector space. Of course vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ can be considered as arrows in each of these spaces. But the properties possessed by these arrows can be generalized to what is known as a vector space. Although the results in this section pertain to all vector spaces, we will see that there is one particular example in field theory that we are interested in.

3.4.1 Basic Facts About Vector Spaces

Definition:

A *vector space* over a field F is an abelian group V with binary operation $+$, with an operation known as scalar multiplication which assigns to each $c \in F$ and $\alpha \in V$ an element $c\alpha \in V$ so that:

1. $r(s\alpha) = (rs)\alpha$ for $r, s \in F$ and $\alpha \in V$.
2. $(r+s)\alpha = r\alpha + s\alpha$ for for $r, s \in F$ and $\alpha \in V$.
3. $r(\alpha + \beta) = r\alpha + r\beta$ for $r \in F$ and $\alpha, \beta \in V$.
4. $1\alpha = \alpha$ for all $\alpha \in V$.

□

Definition:

Let V be a vector space over the field F. A finite set $S = \{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ is *linearly dependent* over F if there is a nontrivial solution $c_1, c_2, \cdots, c_n$ not all zero such that $c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n = 0$. If the only solution to this equation is $c_1 = c_2 = \cdots = c_n = 0$ then the set S is *linearly independent* over F.

□

Example:

The following illustrate linear independence:

- Real Euclidean space $\mathbb{R}^3$ is a vector space over $\mathbb{R}$. The vectors $[0,1,1], [1,0,1], [1,1,1]$ are linearly indepedent over $\mathbb{R}$ since the only real numbers $c_i \in \mathbb{R}$ that satisfy $c_1[0,1,1] + c_2[1,0,1] + c_3[1,1,1] = [0,0,0]$ are $c_1 = c_2 = c_3 = 0$.
- Real Euclidean space $\mathbb{R}^2$ is a vector space over $\mathbb{R}$. The vectors $[-1,1], [0,2], [4,3]$ are not linearly independent over $\mathbb{R}$ because $-8[1,-1] - 7[0,2] + 2[4,3] = [0,0]$

□

Definition:

The set $S = \{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ is a *spanning set of V over F* if every $\alpha \in V$ is a linear combination of the elements of S, i.e. there exist $c_1, c_2, \cdots c_n \in F$ such that

$$\alpha = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n.$$

□

Example:

The following illustrate the concept of spanning:

- Real Euclidean space $\mathbb{R}^2$ is a vector space over $\mathbb{R}$. The vectors $[-1,1], [0,2], [4,3]$ span $\mathbb{R}^2$ over $\mathbb{R}$ because for any $[a,b] \in \mathbb{R}^2$ we can write $[a,b] = a[1,-1] + \frac{a+b}{2}[0,2] + 0[4,3]$
- The set of 2×2 matrices with real entries is a vector space over $\mathbb{R}$, which we will denote $M_2(\mathbb{R})$. The set of matrices $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ does not span $M_2(\mathbb{R})$ as there is no way to express the matrix $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ as a linear combintation of the elements in the set S.

□

A spanning set tells us what vectors we can express in our vector space and linear independence tells us that we are efficiently using the vectors that we have. We can put the idea of linear independence together with a spanning set to form a special sort of spanning set.

Definition:

V is a *finite dimensional vector space over F* if there is a finite spanning set for V over *F*.

□

Definition:

Let V be a vector space over *F*. The set *B* is a *basis* for *V* over *F* if:

1. The set *B* is linearly independent over *F*, and
2. *B* spans V over *F*.

□

Theorem 67. *Let B be a basis for the vector space V over F. Every element of V can be written uniquely as a linear combination of the elements in B.*

Proof. We order the basis elements of *B* as $\alpha_1, \alpha_2, \cdots \alpha_k$ and suppose by way of contradiction that the vector α can be expressed as two different linear combinations of the basis elements:

$$c_1\alpha_1 + c_2\alpha_2 + \cdots + c_k\alpha_k = \alpha = b_1\alpha_1 + b_2\alpha_2 + \cdots + b_k\alpha_k$$

Where there is at least one i such that $c_i \neq b_i$. We rewrite the above equation:

$$(b_1 - c_1)\alpha_1 + (b_2 - c_2)\alpha_2 + \cdots + (b_k - c_k)\alpha_k = 0$$

Since the basis *B* is linearly independent, this means that $b_i - c_i = 0$ for all $i, 1 \leq i \leq k$ and $b_i = c_i$ for all $i, 1 \leq i \leq k$.

□

Although every element can be expressed uniquely in terms of a basis *B*, a basis itself is not unique. For any vector space there can be several different sets of linearly independent spanning sets. What is unique when dealing with different bases is the number of elements in a basis.

Theorem 68. *Let V be a finite dimensional vector space over the field F. Every basis for V over F is a finite set. The number of vectors in any basis of V over F is the same.*

Proof. Let $A = \{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ and $B = \{\beta_1, \beta_2, \cdots, \beta_n\}$ be two bases for *V* over *F*. Since $\alpha_1 \in V$, we may express this as a linear combination of the basis vectors *B*: $\alpha_1 = c_1\beta_1 + c_2\beta_2 + \cdots + c_n\beta_n$

We may write $\beta_1 = c_1^{-1}[\alpha_1 - c_2\beta_2 - \cdots - c_n\beta_n]$. Thus the set $C_1 = \{\alpha_1, \beta_2, \cdots, \beta_n\}$ is a spanning set for *V* over *F*. This set C_1 is also linearly independent. Suppose that $0 = d_1\alpha_1 + d_2\beta_2 + \cdots + d_n\beta_n\}$

□

Definition:

The number of vectors in a basis of the vector space V over F is called the *dimension* of V over F and is denoted $[V : F]$

□

3.4.2 Exercises

1. Let S denote the set of 3×3 symmetric matrices.
   a) Prove that S is a vector space over $\mathbb{R}$.
   b) Find a basis for S.
   c) State the dimension of S.
2. Give a basis for $\mathbb{Q}(\sqrt{6})$ over $\mathbb{Q}$.
3. Give a basis for $\mathbb{Q}(\sqrt[4]{6})$ over $\mathbb{Q}$.
4. Give a basis for $\mathbb{C}$ over $\mathbb{R}$.
5. Give a basis for $\mathbb{Q}(\sqrt[3]{7})$ over $\mathbb{R}$.
6. Give a basis for $\mathbb{R}(i)$ over $\mathbb{R}$.
7. Give a basis for $\mathbb{Q}(\pi)$ over $\mathbb{Q}$.
8. Give a basis for $\mathbb{R}(\pi)$ over $\mathbb{R}$.

9. Let F be a field and let F^n denote the set of ordered n -tuples of elements of *F*. (So $F^2 = \{(a_1, a_2) \mid a_1, a_2 \in F.\}$). Define addition

$$(a_1, a_2, \cdots, a_n) + (b_1, b_2, \cdots, b_n) = (a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n)$$

and define scalar multiplication for $c \in F$:

$$c(a_1, a_2, \cdots, a_n) = (ca_1, ca_2, \cdots, ca_n).$$

Prove that F^n is a vector space over *F*.

10. Let *V* be any finite dimensional vector space over the field *F* of dimension n. Prove that there is a field isomorphism between *V* and F^n from the previous problem.

3.5 Extension Fields

In group theory we started with a group *G* and then determined the subgroup structure of our group. So we were curious about the internal workings of an algebraic object. In field theory the overall philosophy is reversed. We start with a field and then see what other fields we can build upon this one. The focus is upon what we can externally add to a given field, and still have a field.

Definition:

If *E* is an *extension field* of the field *F* if *F* is a subfield of *E*. That is, $F \subseteq E$ and *E* is itself a field.

□

Theorem 69. *Let E be an extension field of the field F, then* E *is a vector space over F.*

Proof. Let $\alpha \in E$. By definition of being a field, *E* is an abelian group. For any $c \in F$ and $\alpha \in E$, $c\alpha \in E$ by the multiplication binary operation in *E*. The other conditions of the scalar multiplication of *E* follow from the fact that of *E* is a field.

□

Definition: If the extension field *E* of *F* is a finite dimensional vector space over *F*, then *E* is a *finite extension* of *F*. The *degree* of *E* over *F* is the dimension of *E* over *F*, which we denote by $[E : F]$.

□

It is no mistake that our term "degree" is doing double duty. Not only is degree applicable in the sense above, but we also saw how this term is used to refer to the highest nonzero term of a polynomial. We will see that there is a connection between these uses of the word degree. We need not stop with one field extension. It is entirely possible to form a sequence of extension fields $F_1 \subseteq F_2 \subseteq \cdots F_n$. We will see an important result dealing with the relative degrees of a sequence of finite field extensions.

Definition:

A sequence of extension fields $F_1 \subseteq F_2 \subseteq \cdots F_n$ is called a *tower* of fields with F_1 the *base field.*

□

One way of forming such a tower of extension fields is to begin with extension field *E* over *F*, and then keep adjoining one element at a time of *E* to *F*.

Definition:

Let *E* be an extension field over *F* and $\alpha \in E$. The field $F(\alpha)$, which is the smallest subfield of *E* with both the elements of *F* and α is a *simple extension of F.*

□

Example:
The field $\mathbb{Q}(\sqrt{2})$ contains elements of the form $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. We see that for any $a + b\sqrt{2} \neq 0 + 0\sqrt{2}$ we have multiplicative inverse $\dfrac{a - b\sqrt{2}}{a^2 - 2b^2} \in \mathbb{Q}(\sqrt{2})$.

□

Example:

Determine the degree of $\mathbb{Q}(\sqrt{5}) = \{a + b\sqrt{5} \mid a, b \in \mathbb{Q}\}$ over the field $\mathbb{Q}$.

We claim that $\{1, \sqrt{5}\}$ is a basis of $\mathbb{Q}(\sqrt{5})$. It is clear that $\{1\}$ is not a basis, as $\sqrt{5} \notin \mathbb{Q}$. The set $\{1, \sqrt{5}\}$ spans $\mathbb{Q}(\sqrt{5})$. Since the basis has dimension two, we say that the degree of $\mathbb{Q}(\sqrt{5})$ over $\mathbb{Q}$ is two and write $[\mathbb{Q}(\sqrt{5}) : \mathbb{Q}] = 2$

□

Theorem 70. *Let D be a finite extension field of E and let E be a finite extension field of F. Then D is a finite extension of F and the degrees are related by $[D : F] = [D : E][E : F]$.*

Proof. We begin by considering a basis for *E* over *F* given by: $\{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ and basis $\{\beta_1, \beta_2, \cdots, \beta_n\}$ for *D* over *E*. The goal will be to show that the set with mn elements

$$A = \{\alpha_i \beta_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$$

is a basis of *D* over *F*. To this end we must show that A spans D over *F* and is linearly independent.

Given an element $d \in D$ we can write $d = d_1\beta_1 + d_2\beta_2 + \cdots + d_n\beta_n$, with $d_i \in E$. We can in turn express each of the d_i as linear combinations $d_i = c_{i1}\alpha_1 + c_{i2}\alpha_2 + \cdots + c_{im}\alpha_m$ with $c_{ij} \in F$. By substitution we see that $d = \sum_{i=1,j=1}^{n,m} c_{ij}\alpha_j\beta_i$, with $c_{ij} \in F$. Thus the set A spans D over *F*.

Now suppose that there are $c_{ij} \in F$ such that $\sum_{i=1,j=1}^{n,m} c_{ij}\alpha_j\beta_i = 0$. Considering this as a linear combination of the β_i, since $\{\beta_i\}$ forms a basis it is linearly independent. So the coefficients of each β_i, $\sum_{j=1}^{m} c_{ij}\alpha_j = 0$. By the fact that the $\alpha)j$ form a basis these too are linearly independent. So $c_{ij} = 0$, showing that A is linearly independent.

We have demonstrated that A is a basis. Because there are mn elements in A we have the expression between the degrees of these extensions: $[D : F] = [D : E][E : F]$.

□

We wrap up this section with a very important application of field extensions. We see that the polynomial $f(x) = x^2 + 4$ has no real zeros, but if we allow ourselves to work in the set of complex numbers there are zeros. In a similar way, the polynomial $x^2 + x + 1 \in \mathbb{Z}_2[x]$ is irreducible over $\mathbb{Z}_2$ and has no zeros. However, there is an extension field E over $\mathbb{Z}_2$ where there is a zero of $x^2 + x + 1$. Kronecker's theorem has a constructive proof in that we don't just show the existence of such an extension field, we also see how to construct such a field.

Theorem 71 (Kronecker's Theorem). *For the field F and nonconstant polynomial $f(x) \in F[x]$ there exists an extension field E of F and $\alpha \in E$ such that $f(\alpha) = 0$.*

Proof. If the polynomial has a zero in *F* then we are done as *F* is an extension field of itself. Otherwise we suppose that f has no zero in F. Let p be an irreducible factor of f. We set $E = F[x]/(p(x))$. By theorem 66 this is a field. We claim this is an extension field of *F* that satisfies the conditions of the theorem.

Let $\phi : F \to F[x]/(p(x))$ be defined by $\phi(a) = \overline{a}$. If $\phi(a) = \phi(b)$ then $\overline{a} = \overline{b}$ and $(a - b) = g(x)p(x)$. Since the degree of $p(x)$ is greater than zero, $g(x) = 0$ and $a - b = 0$, thus $a = b$. By the definition of our addition and multiplication in $F[x]/(p(x))$, ϕ is a homomorphism of fields that maps *F* into E. Thus E is an extension field of *F*.

Now consider $\alpha = \overline{x} \in E$. We see that

$$p(\alpha) = a_n\overline{x}^n + a_{n-1}\overline{x}^{n-1} + \cdots + a_1\overline{x} + a_0 = \overline{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0} = \overline{p(x)} = \overline{0},$$

□

3.5.1 Exercises

1. Let E be a finite extension of F and suppose that $[E : F]$ is prime. Prove that E is a simple extension of *F*.
2. Form two towers of fields with base $\mathbb{Q}$ and extension field $\mathbb{Q}(\sqrt{7}, \sqrt[3]{2})$ at the top.
3. Given that E is an extension field of the field *F*, prove that $[E : F] = 1$ if and only if $E = F$.
4. Given that E is an extension field of the field *F*, prove that if $[E : F]$ is prime then there is no field K such that $F \subseteq K \subseteq E$.
5. Given that $\alpha \in E$ has degree n over *F*, prove that n divides $[E : F]$.

3.6 Algebraic Extensions

We start by considering the field extension $\mathbb{Q} \subseteq \mathbb{R}$. Since $\mathbb{Q}$ is a subfield of $\mathbb{R}$ this is a field extension. There are different sorts of elements in $\mathbb{R}$ as we consider them over $\mathbb{Q}$. For instance, the element $\sqrt{2} \notin \mathbb{Q}$, but yet we can express $\sqrt{2}$ as the zero of a polynomial $f(x) \in \mathbb{Q}[x]$. One such polynomial is $f(x) = x^2 - 2$. On the other hand, there is no polynomial with coefficients in $\mathbb{Q}$ with the number π as a zero. We wish to sort out these ideas, and to connect them with what we have learned about polynomials and vector spaces.

Definition: Let *E* be an extension field of the field *F*.

- If $\alpha \in E$ is the zero of some polynomial over *F* we say that α is *algebraic over F*.
- If every element $\alpha \in E$ is algebraic over *F* then we say that E is an *algebraic extension of* F.
- If $\alpha \in E$ is not algebraic over *F* we say that it is *transcendental over F*.

□

Example: The discussion above shows that $\sqrt{k} \in \mathbb{C}$ for any $k \in \mathbb{Z}$ is algebraic over $\mathbb{Q}$. The proof is beyond the scope of this book, but $\pi \in \mathbb{C}$ is transcendental over $\mathbb{Q}$.

□

Example: For any field *F* and field extension E all of the elements of *F* itself are algebraic over *F*. We see this because if $a \in F$ then a is a zero of $f(x) = (x - a) \in F[x]$.

□

Definition: Let E be an extension field of *F* and let $\alpha \in E$ be algebraic over *F*. Form the set

$$M = \{f \in F[x] \mid f(\alpha) = 0\}.$$

Choose an element g of M of minimal degree. The polynomial $g(x)$ is a *minimal polynomial of* α *over* F

□

There are two properties of minimal polynomials that connect them to previous topics, and will be important as we continue.

Theorem 72. *If* g *is a minimal polynomial for* α *over* F *then* g *is irreducible over F.*

Proof. Suppose by way of contradiction that g is reducible. There exist $h, k \in F[x]$ with degree less than g such that $g(x) = h(x)k(x)$. Since $g(\alpha) = 0$ we have $h(\alpha)k(\alpha) = 0$ and so either $h(\alpha) = 0$ or $k(\alpha) = 0$. In either case we have a contradiction to the fact that g is of minimal degree with α as a zero. Thus g is irreducible.

□

Theorem 73. *If* g *is a minimal polynomial for* α *over F and* $h(\alpha) = 0$ *then then* g *divides* h.

Proof. Suppose that $h(\alpha) = 0$. We use the division algorithm and see that $h(x) = g(x)q(x) + r(x)$. We then see that

$$0 = h(\alpha) = g(\alpha)q(\alpha) + r(\alpha) = 0 + r(\alpha) = r(\alpha)$$

If the degree of r is less than the degree of g then this contradicts the fact that g is of minimal degree with α as a zero. Thus $r(x) = 0$ and $h(x) = g(x)q(x)$.

□

It is a very easy consequence of these preceeding theorems that if g, f are both minimal polynomials of α over *F* then $f = cg$ where $c \in F$.

Example:

What is a minimal polynomial for $\sqrt{3}$ over $\mathbb{Q}$?

Here we see $\sqrt{3}$ is a zero of $x^2 - 3 \in \mathbb{Q}[x]$. This polynomial is of minimal degree because for if there was a polynomial of degree 1 with $\sqrt{3}$ as a zero, it would imply that $\sqrt{3} \in \mathbb{Q}$.

□

Example:

What is a minimal polynomial for $\alpha = \sqrt{3} + \sqrt{5}$ over $\mathbb{Q}$?

We begin by noting that there is no polynomial of degree 1 over $\mathbb{Q}$ such that $f(\alpha) = 0$. By squaring α we see: $\alpha^2 = 3 + 2\sqrt{3}\sqrt{5} + 5 \Rightarrow \alpha^2 - 8 = 2\sqrt{3}\sqrt{5}$. We now square both sides of this equation and obtain $(\alpha^2 - 8)^2 = 4 \cdot 3 \cdot 5 \Rightarrow \alpha^4 - 16\alpha^2 + 64 = 60$. Therefore α is a zero of the polynomial $f(x) = x^4 - 16x^2 + 4$.

f has no rational zeros as the only candidates, the factors of constant term 4, do not work. Thus f is irreducible and minimal.

□

Example:

What is a minimal polynomial for $\alpha = \sqrt{3} + \sqrt{5}$ over $\mathbb{Q}(\sqrt{5})$?

Here the situation is different than the preceeding example. We are allowed to use $\sqrt{5}$ for the coefficients of our minimal polynomial. Again we see that $\alpha^2 = 3 + 2\sqrt{3}\sqrt{5} + 5 \Rightarrow 0 = \alpha^2 - 2\sqrt{3}\sqrt{5} - 8$, however we also see that $\sqrt{5}\alpha = \sqrt{5}(\sqrt{3} + \sqrt{5} = \sqrt{3}\sqrt{5} + 5$ and thus $\sqrt{3}\sqrt{5} = \sqrt{5}\alpha - 5$. Substituting this into our expression we have:

$$0 = \alpha^2 - 2(\sqrt{5}\alpha - 5) - 8 = \alpha^2 - 2\sqrt{5}\alpha + 2 \in (\sqrt{5})[x].$$

Therefore a minimal polynomial for $\sqrt{3} + \sqrt{5}$ over $\mathbb{Q}(\sqrt{5})$ is $f(x) = x^2 - 2\sqrt{5}x + 2$.

□

Definition:

Let E be an extension field of F and $\alpha \in E$ algebraic over *F*. The simple extension $F(\alpha)$ is said to be a *finite extension of degree* n where n is the degree of a minimal polynomial of α over *F*.

□

Theorem 74. *Let F be a field and E an extension field of F. If $\alpha \in E$ is algebraic over F of degree n then $F(\alpha)$ is a vector space over F with basis $\{1, \alpha, \alpha^2, \cdots, \alpha^{n-1}\}$*

Proof. We know that $F(\alpha)$ is a field containing the elements $\{1, \alpha, \alpha^2, \cdots, \alpha^{n-1}\}$ as well as every linear combination of these elements: $c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{n-1}\alpha^{n-1}$ where $c_i \in F$. It is clear that this set of linear combinations is spanned by $\{1, \alpha, \alpha^2, \cdots, \alpha^{n-1}\}$. If we suppose that $c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{n-1}\alpha^{n-1} = 0$, with at least one $c_i \neq 0$ then we have a polynomial of degree $n-1$ of which α is a zero. This contradicts the fact that α is of degree n. Thus $c_i = 0$ for $1 \leq i \leq n-1$ and the set $\{1, \alpha, \alpha^2, \cdots, \alpha^{n-1}\}$ is linearly independent over *F*.

□

Example:
We have seen that $\sqrt{3} + \sqrt{5}$ is algebraic of degree 4 over $\mathbb{Q}$. The set $\{1, \sqrt{3} + \sqrt{5}, (\sqrt{3} + \sqrt{5})^2, (\sqrt{3} + \sqrt{5})^3\}$ forms a basis for $\mathbb{Q}(\sqrt{3} + \sqrt{5})$ over $\mathbb{Q}$.

Example:

Let ζ be a pth root of unity. Recall that these are the complex numbers which are zeros of the polynomial $f(x) = x^p - 1$. We have see that $f(x) = (x-1)\Phi_p(x)$ where $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x^2 + x + 1$ and that $\Phi_p(x)$ is irreducible over $\mathbb{Q}$. It follows that $\Phi_p(x)$ is a minimal polynomial for ζ. By theorem 74 $\mathbb{Q}(\zeta)$ has degree $p-1$ over $\mathbb{Q}$ and $\{1, \zeta, \zeta^2, \zeta^3, \cdots, \zeta^{p-2}\}$ is a basis for $\mathbb{Q}(\zeta)$ over $\mathbb{Q}$.

□

3.6.1 Theorems Regarding Algebraic Extensions

Algebraic extensions are important because they allow us to classify other extensions.

Theorem 75. *Any finite extension is also an algebraic extension.*

Proof. Let E be a finite extension of F with degree of n. Choose any element $\alpha \in E$ and we know that $\{1, \alpha, \alpha^2, \cdots, \alpha^{n-1}, \alpha^n\}$ is linearly dependent. So there exist $c_i \in F$ that are not all equal to zero such that $c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_n\alpha^n = 0$. The element $\alpha \in E$ is a zero of a polynomial $c_0 + c_1x + c_2x^2 + \cdots + c_nx^n \in F[x]$. Therefore α is algebraic over *F*.

□

By this theorem we know that if α is algebraic over a field *F*, then the field $F(\alpha)$ is algebraic. This result can be extended to a tower of fields.

Theorem 76. *Let $F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n$ be a tower of fields. If $F_i = F_{i-1}(\alpha_i)$ for $1 \leq i \leq n$ and elements α_i algebraic over F_{i-1}, then F_n is algebraic over F_0.*

Proof. Since each extension is finite, by theorem 70 F_k is a finite extension over F_0. By theorem 75 F_k is also an algebraic extension over F_0.

□

Theorem 77. *If E is an algebraic extension of F and D is an algebraic extension of E then D is an algebraic extension of F.*

Proof. Begin with an element $\alpha \in D$ with minimal polynomial with coefficients $d_0, d_1, \cdots, d_n$. We have tower of fields $F \subseteq F(d_0) \subseteq F(d_0, d_1) \subseteq \cdots \subseteq F(d_0, d_1, \cdots, d_n)$. The element α is algebraic over the field $F(d_0, d_1, \cdots, d_n)$. We add one more level to the tower: $F(d_0, d_1, \cdots, d_n, \alpha)$ and by theorem 76 see that α is algebraic over *F*.

□

3.6.2 Finite Fields

We have already seen that for any prime number p, $\mathbb{Z}_p$ with addition and multiplication modulo p forms a field. Kronecker's theorem and theorem 74 allow us to construct finite fields of order p^k where p is a prime.

1. Begin with $\mathbb{Z}_p$
2. Use any irreducible polynomial $f \in \mathbb{Z}_p[x]$ of degree k.
3. Form the extension field $E = \mathbb{Z}_p[x]/(f)$.
4. Let α be the element $\overline{x} \in \mathbb{Z}_p[x]/(f)$.

5. The p^k elements of E are $c_{k-1}\alpha^{k-1} + c_{k-2}\alpha^{k-2} + \cdots + c_1\alpha + c_0$ with $c_i \in \mathbb{Z}_p$.
6. The multiplicative structure of E is governed by the polynomial f. We may express α^k in terms of the basis $\{1, \alpha, \alpha^2, \cdots, \alpha^{k-1}\}$.

Example:

Construct a field with 8 elements.

Since $8 = 2^3$ we need to start with an irreducible polynomial of degree 3 in $\mathbb{Z}_2[x]$. One such polynomial is $f(x) = x^3 + x + 1$. In $E = \mathbb{Z}_2[x]/(x^3 + x + 1)$ we let $\alpha = \overline{x}$. There are elements $0, 1, \alpha, \alpha^2, \alpha + 1, \alpha^2 + 1, \alpha^2 + \alpha, \alpha^2 + \alpha + 1$. By the polynomial $f(x) = x^3 + x + 1$ we have $\alpha^3 + \alpha + 1 = 0$, or $\alpha^3 = \alpha + 1$.

Once we have this relationship, and remember that $1 + 1 = 0$ in $\mathbb{Z}_2$, calculations are straightforward:

$$(\alpha^2 + \alpha + 1) \cdot (\alpha + 1) = \alpha^3 + \alpha^2 + \alpha + \alpha^2 + \alpha + 1 = \alpha^3 + 1 = \alpha + 1 + 1 = \alpha$$

□

3.6.3 Exercises

1. Construct a field with nine elements, showing the addition and multiplication tables.
2. Find the degree of the extension $\mathbb{Q}(\sqrt{7})$ over $\mathbb{Q}$.

3. Find the degree of the extension $\mathbb{Q}\left(\sqrt{\sqrt{2}+\sqrt{5}}\right)$ over $\mathbb{Q}$.
4. Find the degree of the extension $\mathbb{Q}\left(\sqrt{2},\sqrt{3},\sqrt{6}\right)$ over $\mathbb{Q}$.
5. Find the degree of the extension $\mathbb{Q}\left(\sqrt[3]{7}\right)$ over $\mathbb{Q}$.
6. Find the degree of the extension $\mathbb{Q}\left(\sqrt{3},\sqrt{15}\right)$ over $\mathbb{Q}\left(\sqrt{5}\right)$.
7. Prove that for p,q prime and $p \neq q$ that $\mathbb{Q}\left(\sqrt{p}+\sqrt{q}\right) = \left(\sqrt{p},\sqrt{q}\right)$.
8. Find a minimal polynomial for $\sqrt{2+\sqrt[3]{5}}$ over $\mathbb{Q}$.

3.7 Geometric Constructions

"To bisect a given rectilinear angle. Let the angle BAC be the given rectilinear angle. It is required to bisect it." – Euclid

At long last we come to the stated goal of the introduction, the topic of the possibility of certain geometric constructions. This highlights an unexpected connection between abstract algebra and geometry, and answers questions that the Greeks asked when they first developed the careful study of geometry. It is clear that any angle can be bisected by using a straightedge and compass. In other words, given any angle in the plane, we can use a straightedge and compass to construct an angle with measure exactly half of our original angle. In *The Elements* Euclid demonstrates how to do this early in his textbook. Proposition 9 of book I shows how to bisect any angle.

What is not clear is that any angle can be trisected by using a straightedge and compass. Given any angle, can we use our tools to construct an angle with measure exactly one third of our original angle? Of course, this is true for certain angles. An angle of 180^o can be trisected because it is possible to construct an equilateral triangle, with angle measures of 60^o. But given any angle θ can we construct the angle $\theta/3$? The answer is that this is not possible to do. The reason why is due to applications of our field theory. Before any abstract algebra shows up, we will take a further detour into geometry.

3.7.1 Constructible Numbers

What is really happening when we use a straightedge and compass for a geometric construction? We are using our geometric tools in such a way to produce a line segment of a given length, circle of a particular radius, or angle of a given measure. In each of these cases we arrive at a number, which motivates our next defintion.

Definition: The number θ is a *constructible number* if a line segment of length $|\theta|$ can be constructed in a finite number of steps with a compass and straightedge.

□

Definition: The point (x,y) in the plane is a *constructible point* if it can be constructed in a finite number of steps with a compass and straightedge.

□

We now translate our axioms from geometry into the language of constructible numbers.

1. The points $(0,0)$ and $(1,0)$ are constructible.
2. Two constructible points determine a constructible line segment or line.
3. Any circle with center point constructible and radius a constructible number is constructible.
4. Two constructible lines intersect at a constructible point.
5. A constructible line and constructible circle intersect at constructible point(s).
6. Two constructible circles intersect at constructible point(s).

Theorem 78. *The integers are constructible numbers.*

Proof. We start with our line segment with endpoints $(0,0),(1,0)$. This has unit length and can be extended indefintely using our straightedge. Our compass can transfer the length 1 a total of k times, where k is a positive integer. Thus the set $\mathbb{Z}$ is constructible.

□

Theorem 79. *If θ and η are constructible real numbers, then $\theta+\eta, \theta-\eta$, and $\theta\cdot\eta$ are also constructible. If θ and η are constructible real numbers, and $\eta\neq 0$ then θ/η is also constructible.*

Proof. Suppose that θ and η are constructible. It is clear that $\theta+\eta$ is constructible, since given the lengths $|\theta|$ and $|\eta|$ a line segments of lengths $|\theta|+|\eta|$ and $||\theta|-|\eta||$ are constructible. Furthermore since $|-\eta|=|\eta|$ if η is constructible then so is $-\eta$, thus $\theta-\eta$ is constructible.

We construct $|\theta\eta|$ as follows:

1. Start with line segment of length $|\theta|$ with endpoints 0 and *P*
2. Form a ray by extending $\vec{OP}$ indefinitely in direction of *P*.
3. Construct a line segment of length 1 with one endpoint 0 and the other *Q* not on the ray $\vec{OP}$
4. Form a ray by extending $\overline{OQ}$ indefinitely in direction of *Q*
5. Construct η on $\vec{OQ}$ with endpoints 0 and *R*.
6. Construct line segment $\overline{PQ}$
7. Construct a line parallel to $\overline{PQ}$ through the point *R*. Label its intersection with $\vec{OP}$ by *S*.

We now have similar triangles ΔOQP and ΔORS. Since corresponding sides are in proportion we have:

$$\frac{OQ}{OP}=\frac{OR}{OS}\Rightarrow\frac{1}{|\theta|}=\frac{|\eta|}{OS}\Rightarrow OS=|\theta\eta|$$

and so $\theta\eta$ is constructible.

Now we suppose that $\eta \neq 0$. To construct $\left|\frac{1}{\eta}\right|$ We follow a similar construction as above:

1. Start with line segment of length 1 with endpoints 0 and P
2. Construct a line segment of length 1 with one endpoint 0 and the other Q not on the line $\overline{OP}$
3. Form a ray by extending $\overline{OQ}$ indefinitely in direction of Q
4. Construct η on $\vec{OQ}$ with endpoints 0 and R.
5. Construct line segment $\overline{RP}$
6. Construct a line parallel to $\overline{RP}$ through the point Q. Label its intersection with $\vec{OP}$ by S.

We now have similar triangles ΔOQS and ΔORP. Since corresponding sides are in proportion we have:

$$\frac{OS}{OQ} = \frac{OP}{OR} \Rightarrow \frac{OS}{1} = \frac{1}{|\eta|} \Rightarrow OS = \frac{1}{|\eta|}$$

and so $1/\eta$ is constructible.

It then follows that the product $\theta\frac{1}{\eta} = \frac{\theta}{\eta}$ is constructible.

□

Let us take inventory of our set of constructible numbers $\mathbb{T}$. This set is not just the set of integers, since we can form quotients $\frac{\theta}{\eta}$, rational numbers are also constructible. Since we have the commutative operations of sum, product, and their inverses, the set of constructible numbers forms a field. It is clear that the field of rationals is a subfield of $\mathbb{T}$. Is it also true that $\mathbb{Q} = \mathbb{T}$? A little bit of thought tells us that this is not the case.

Theorem 80. *For all positive integers n, the number $\sqrt{n}$ is constructible.*

Proof. The proof is by induction on n. We begin by noting that $\sqrt{1} = 1$ is constructible.

Now construct an isosceles triangle with side lengths of 1. This is possible because we can draw a line perpendicular to a given line, and then use the compass to mark a length of 1 on both perpendicular lines. The hypotenuse of this triangle has length $\sqrt{1^2 + 1^2} = \sqrt{2}$.

By our induction hypothesis $\sqrt{k}$ is constructible. Now construct a right triangle with legs of length 1 and $\sqrt{k}$. The hypotenuse has length $\sqrt{k+1}$.

□

This shows that the set of constructible numbers contains more than just the rational numbers. The question becomes, how much more? We go back to our list of axioms for constructible points and figures in the plane. It is clear that any points (x, y) in the plane where x, y are rational numbers are constructible. Given two pairs of points with rational coordinates, the intersection of the lines formed by each pair is constructible. However, this intersection point (if it exists) will result in another point with rational coordinates. In other words, we do not get any new constructible points from the intersection of two constructible lines of this type.

If two circles intersect, then they do so at a single point or at two points. In either of these cases we can obtain the same intersection points by the intersection of a circle with a line. Thus the only remaining item of consideration is the intersection of a constructible line that passes through a pair of points with rational coordinates with a constructible circle. This circle will have rational radius and a center with rational coordinates. We will determine at the intersection points of a circle $(x-h)^2 + (y-k)^2 = r^2$ with the line $y = mx + b$ where $m, b, h, k, r \in$. These intersection points are constructible. We see that constructible points are found by solving the equation

$$(x-h)^2 + (mx + b - k)^2 = r^2 \Rightarrow$$

$$(x^2 + m^2x^2) - 2xh + 2m(b-k)x + (b-k)^2 + h^2 - r^2 = 0.$$

This is a quadratic, and so the solutions include $x = \sqrt{a}$ for $a \in \mathbb{Q}$.

Of course we could continue this process and use $\sqrt{q}$ for $q \in \mathbb{Q}$ as the coordinates. Since all that we do in solving a quadratic is to use field operations and a square root, this suggests the following theorem.

Theorem 81. *If $\theta > 0$ then $\sqrt{\theta}$ is constructible.*

Proof. The line $y = \frac{\theta}{4} - 1$ is constructible, as is the circle with center $(0, 0)$ and radius $\frac{\theta}{4} + 1$. This circle has equation $x^2 + y^2 = \left(\frac{\theta}{4} + 1\right)^2$. The intersection points of this circle with the line are also constructible. These points satisfy the equation of the line and of the circle: $x^2 + \left(\frac{\theta}{4} - 1\right)^2 = \left(\frac{\theta}{4} + 1\right)^2$

$$x^2 + \frac{\theta^2}{16} - \frac{\theta}{2} + 1 = \frac{\theta^2}{16} + \frac{\theta}{2} + 1 \Rightarrow x^2 = \theta.$$

Thus $x = \sqrt{\theta}$ and so $\sqrt{\theta}$ is constructible.

□

The above discussion shows that constructible numbers are produced by starting with a rational number and then applying a finite number of square roots and field operations. Combining theorem 81 and theorem 79 we have the following theorem:

Theorem 82. *The field of constructible numbers consists of all real numbers obtained from $\mathbb{Q}$ by applying a finite number of square roots of positive numbers and field operations.*

Theorem 83. *If θ is a constructible number, then $[\mathbb{Q}(\theta) : \mathbb{Q}] = 2^k$ for some nonnegative integer k.*

Proof. Let θ be a constructible number. By theorem 82 there are $a_1, a_2, \cdots, a_n \in \mathbb{R}$ such that $\mathbb{Q}(a_1, a_2, \cdots, a_j)$ is a degree 2 extension of $\mathbb{Q}(a_1, a_2, \cdots, a_{j-1})$, and that $\theta \in \mathbb{Q}(a_1, a_2, \cdots, a_n)$. Thus we have

$$2^n = [\mathbb{Q}(a_1, a_2, \cdots, a_n) : \mathbb{Q}] = [\mathbb{Q}(a_1, a_2, \cdots, a_n) : \mathbb{Q}(\theta)][\mathbb{Q}(\theta) : \mathbb{Q}],$$

and therefore $[\mathbb{Q}(\theta) : \mathbb{Q}] = 2^k$ for some $k \geq 0$.

□

Theorem 83 is quite powerful as it determines the possibility of a construction without actually performing the construction. We now more or less effortlessly can state that it is impossible to trisect a given angle. First we note the following:

Theorem 84. *The angle θ is constructible if and only if $|\cos\theta|$ is constructible.*

Proof. Suppose θ is constructible. Construct a right triangle with hypotenuse of length 1 with angle θ. By basic trigonometry the side adjacent to θ has length $|\cos\theta|$

Now suppose that $|\cos\theta|$ is constructible. Construct a right triangle with hypotenuse of length 1 and one leg of length $|\cos\theta|$. By trigonometry the angle adjacent to the side of length $|\cos\theta|$ is θ.

□

An angle of 60^o can be constructed as this is the measure of each of the angles in an equilateral triangle. We will show that it is impossible to trisect this angle. Since it is impossible to trisect an angle of measure 60^o it is impossible, in general, to trisect a given angle.

Theorem 85. *The angle 60^o cannot be trisected by a compass and straightedge.*

Proof. By theorem 84 we will show that $\cos 20^o$ is not constructible. We use some trigonometric identities, including $\cos(\alpha+\beta), \cos 2\alpha, \sin 2\alpha, 1 = \sin^2\alpha + \cos^2\alpha$:

$$\begin{aligned}
\cos 3\theta &= \cos(\theta + 2\theta) \\
&= \cos\theta\cos 2\theta - \sin\theta\sin 2\theta \\
&= \cos\theta(2\cos^2\theta - 1) - 2\sin\theta\cos\theta\sin\theta \\
&= \cos\theta(2\cos^2\theta - 1) - 2\cos\theta(1 - \cos^2\theta) \\
&= 4\cos^3\theta - 3\cos\theta
\end{aligned}$$

Thus if $\theta = 20^o$ then $\cos 60^o = 4(\cos 20^o)^3 - 3(\cos 20^o)$ and we have

$$4(\cos 20^o)^3 - 3(\cos 20^o) - \frac{1}{2} = 0.$$

This polynomial is irreducible over $\mathbb{Q}$ and so the degree of the minimal polynomial of $\cos 20^o$ over $\mathbb{Q}$ is 3. In order for $\cos 20^o$ to be constructible we would need $3 = [\mathbb{Q}(\cos 20^o) : \mathbb{Q}] = 2^k$ for some $k \geq 0$. This is impossible, so $\cos 20^o$ and thus an angle of 20^o is not constructible.

□

3.7.2 Exercises

1. Prove that it is not possible with straightedge and compass to construct a square with the same area as a given circle.
2. Prove that it is not possible with straightedge and compass to construct a cube with volume double that of a given cube.
3. Show using abstract algebra that it is possible to trisect a 90º angle.

4 Ring Theory

4.1 Introduction to Rings

Now that we have spent some time studying fields, we will take a brief excursion to a more general setting. Like a field, this new algebraic structure again has two operations, which we will call addition and multiplication. The example that it will be helpful to think about as we continue is that of the integers $\mathbb{Z}$. There are two binary operations associated with the integers, that of addition and multiplication. The set of integers under addition is an abelian group. As we have seen, this set is not a group under multiplication. Nearly all of the elements in $\mathbb{Z}$ do not have a multiplicative inverse. There is no integer z such that $2z = 1$. The multiplication is associative, and interacts with addition by means of the distributive property. The multiplicative structure of our structure that mimics the integers is much more relaxed than that of a field. Indeed, the multiplication need not even be commutative. This structure is known as a ring.

Definition:

A *ring* is a set R with two binary operations, which we will call addition +, and multiplication $\cdot$ subject to the following conditions:

1. R with the addition operation is an abelian group.
2. Multiplication in R is associative.
3. Multiplication and addition interact with each other according to the distributive properties. For all $a, b, c \in R$:
   a) $a \cdot (b + c) = a \cdot b + a \cdot c$
   b) $(a + b) \cdot c = a \cdot c + b \cdot c$

□

It is important with a definition such as the one above to notice what is *not* stated. The multiplication operation does not need to be commutative. There is also no mention of multiplicative inverses or a multiplicative identity. We can add to our definition of ring to include these features.

Definition:

A *commutative ring* is a ring for which the multiplication operation is commutative.

□

Definition:

A *ring with unity* is a ring for which there is a multiplicative identity.

□

Example:

We have already mentioned the set of integers $\mathbb{Z}$ as an example of a ring. Given the further definitions above, to be precise $\mathbb{Z}$ is a commutative ring with unity.

□

Example:

The set of rational numbers $\mathbb{Q}$ is a commutative ring with unity under the operations of fraction addition and fraction multiplication:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

It is a very tedious exercise to check that this is a ring:

- We have already seen that under fraction addition, the rational numbers form an abelian group.
- We now check for associativity of multiplication:
  $\left(\frac{a}{b} \cdot \frac{c}{d}\right) \cdot \frac{e}{f} = \frac{(ac)e}{(bd)f} = \frac{a(ce)}{b(df)} = \frac{a}{b} \cdot \left(\frac{c}{d} \cdot \frac{e}{f}\right)$
- The multiplication is commutative: $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} = \frac{ca}{db} = \frac{c}{d} \cdot \frac{a}{b}$

- Due to the commutativity of multiplication, we only need to check one distributive properties $\frac{a}{b} \cdot \left(\frac{c}{d} + \cdot \frac{e}{f} \right) = \frac{a}{b} \cdot \left(\frac{cf + de}{df} \right) = \frac{a(cf + de)}{b(df)} = \frac{a(cf) + a(de)}{b(df)}$

Unity here is the rational number 1/1, since $\frac{a}{b} \cdot \frac{1}{1} = \frac{a}{b}$.

□

Example:

We have already seen that $\mathbb{Z}_6$ is a group under addition modulo 6. If we also include multiplication modulo 6, this set with these two operations is a ring.

$\cdot$	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

There is nothing special about the number 6 above. The set $\mathbb{Z}_n$ under modulo addition and multiplication is a ring for any positive integer n.

□

In the above example, we can see from the multiplication table that we do not have a group table. Some elements, such as 3, do not have a multiplicative inverse. Other elements, such as 5, do have a multiplicative inverse. We signify these types of elements with the following definition.

Definition:

Elements of a ring with a multiplicative inverse are called *units*.

□

WARNING:

Despite the similarity in form and meaning, be sure to note the distinction between unity and a unit. Unity (if it exists) in a ring is the unique element that is a multiplicative identity. A unit is an element that has a multiplicative inverse. *Unity is a unit, but a unit may not be unity.*

□

Example:

Let $[\sqrt{-7}]$ denote the set $\{a+b\sqrt{-7} \mid a,b \in\}$. This set is a commutative ring with unity under addition $(a+b\sqrt{-7})+(c+d\sqrt{-7})=(a+c)+(b+d)\sqrt{-7}$ and multiplication $(a+b\sqrt{-7})\cdot(c+d\sqrt{-7})=(ac-7bd)+(ad+bc)\sqrt{-7}$. (This addition and multiplication could be derived by basic algebra, taking note that $(\sqrt{-7})^2=-7$.

Unity in this ring is simply $1+0\sqrt{-7}$

To determine the units of this ring we ask, for which elements $(a+b\sqrt{-7})$ is there an element $(c+d\sqrt{-7})$ such that $(a+b\sqrt{-7})\cdot(c+d\sqrt{-7})=1+0\sqrt{-7}$?

We perform the multiplication and see that we want $(ac-7bd)+(ad+bc)\sqrt{-7}=1+0\sqrt{-7}$. This results in the equations:

$$ac-7bd=1 \qquad ad+bc=0$$

Since a,b are known constants, we solve for c and d and see that $\begin{array}{rcrcl} abc & - & 7b^2d & = & b \\ -abc & - & a^2d & = & 0 \end{array}$

We add the equations, solve for d and see that $d=\dfrac{-b}{a^2+7b^2}$ and $c=$

□

Example:

For any ring R the set of all polynomials with coefficients in R, which we denote $R[x]$, is a ring under polynomial addition and multiplication. Some of the structure of R is inherited in $R[x]$. For instance, if R is a commutative ring then $R[x]$ is a commutative ring. If R has unity then $R[x]$ also does.

□

Example:

Let $M_2(\mathbb{Z})$ denote 2 by 2 matrices with integers as entries. This is a ring under matrix addition and matrix multiplication. More specifically this is a noncommutative ring with unity:

$$\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 1 & -2 \end{bmatrix}=\begin{bmatrix} 5 & -4 \\ 5 & -2 \end{bmatrix}\neq\begin{bmatrix} 3 & 6 \\ -3 & 4 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$$

The unity in this ring is the identity matrix $I_2=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

There are many units in this ring. For example, $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}^2 = I_2$, so this matrix is its own inverse.

To determine all units of this ring we ask, for what matrices $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is there a matrix B such that $AB = I_2$?

Let $B == \begin{bmatrix} e & f \\ g & h \end{bmatrix}$

From the matrix equation $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} =] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ we obtain the following system of linear equations:

$$ae + cg = 1 \quad af + bh = 0 \quad ce + gd = 0 \quad cf + dh = 1$$

The solution of this system shows that the matrix $B = \frac{1}{ad - bc} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$. The entries of this matrix are integers if and only if $ad - bc = 1$. This means that the units of the ring are matrices for which the condition $ad - bc = 1$ holds.

□

Notation:

The set of units of a ring R is denoted by R^*.

□

Theorem 86. *Given a ring with unity R, the units of this ring R^* form a group under the multiplication operation.*

Proof. Since R has unity and this is a unit, this means that R^* is nonempty. We begin by showing the set of units is closed under multiplication. Given any $x, y \in R^*$ we know that there exists a multiplicative inverse x^{-1}, y^{-1}. These are also units, and are elements of R^*. Since $(xy)(y^{-1}x^{-1}) = 1$, this shows that $(xy)^{-1}$ exists and is a unit. Therefore $xy \in R^*$.

We now check the group axioms. Multiplication in a ring is associative. So R^* inherits this property from R. Unity is an element of R^*, so R^* has an identity element. If $x \in R^*$ then x is a unit. So there is a multiplicative inverse x^{-1}. Since $x \cdot x^{-1} = 1$, it follows that x^{-1} is also a unit. Thus $x^{-1} \in R^*$, and if $x \in R^*$ then $x^{-1} \in R^*$.

□

4.1.1 Further Structure

Analogs of topics that we saw in our study of groups exist in the setting of rings. For example, just as we formed the direct product of groups, it is possible to form a direct product of rings. Homomorphisms and isomorphisms of rings can also be constructed. Not only do these maps respect the addition operation, they also respect the multiplication operation.

Definition:

The mapping $\phi : R \to S$ is a *ring homomorphism* if $\phi(x + y) = \phi(x) + \phi(y)$ and $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in R$.

A ring homomorphism is a *ring isomorphism* if it is one-to-one and onto.

□

Definition: Given the rings R_1, R_2 with addition $+_1, +_2$ and multiplication $\cdot_1, \cdot_2$ respectively, the *direct product of rings* is the Cartesian product $R_1 \times R_2$ with addition $(a_1, a_2) + (b_1, b_2) = (a_1 +_1 b_1, a_2 +_2 b_2)$ and multiplication $(a_1, a_2) \cdot (b_1, b_2) = (a_1 \cdot_1 b_1, a_2 \cdot_2 b_2)$

□

4.1.2 Exercises

1. Let *R* be a commutative ring with unity. What are the units of the ring $R[x]$?
2. What are the units of the ring $\mathbb{Z}_{12}$?
3. What are the units of the ring $\mathbb{Z} \times \mathbb{Z}$?
4. For *F* a field is $F[x]$ a field? Explain.
5. A Boolean ring is a ring *R* such that $x^2 = x$ for all $x \in R$. Prove that a Boolean ring is a commutative ring.

6. Given a set *S* form the set of subsets of *S*, known as the power set $P(S)$. For every $A, B \in P(S)$ we define

$$A + B = A \Delta B \qquad A \cdot B = A \cap B$$

where $A \Delta B$ denotes the symmetric difference of A and *B*. Prove that $P(S)$ with these operations forms a commutative ring.

7. A ring element is *nilpotent* if $a^n = 0$ for some $n \in^+$. Prove that if $a, b \in R$ where R is a commutative ring and a, b are nilpotent then $a + b$ is nilpotent.

4.2 Integral Domains

At times in abstract algebra, we are at a disadvantage because we know too much. The algebra from our younger days where there were a lot more numbers and a lot less proofs ï¿½ is true, but the setting of this algebra was the set of real numbers. If we rely too much upon this algebra, then we assume too much. There are many algebraic features of the real numbers that are actually quite special. We have already seen many of these properties in our study of fields. One such property involves something that is sometimes called the zero product property. We have been taught in algebra that if $xy = 0$ then $x = 0$ or $y = 0$. There is actually a specialized setting that makes this true.

Consider the ring $\mathbb{Z}_{10}$. Under addition modulo 10 and multiplication modulo 10 this is a commutative ring with unity, but not a field. When we examine the multiplicative structure of the ring $\mathbb{Z}_{10}$ we notice something that has a bearing on the above discussion. The elements 2 and 5 are nonzero, however $2 \cdot 5 = 0$. This is not the only instance of this in $\mathbb{Z}_{10}$. The products of nonzero elements $4 \cdot 5 = 6 \cdot 5 = 8 \cdot 5 = 0$.

Definition:

Let $r, s \in R$ be two nonzero elements. If $rs = 0$ then we call r and s *zero divisors.*

□

Theorem 87. *The zero divisors of* n *are all nonzero elements that are not relatively prime to* n.

Proof. Let $r \in_n$ with $r \neq 0$. We suppose that r and n are not relatively prime to each other, that is $\gcd(n, r) = d \neq 1$. We see that $\frac{n}{d} r = \frac{r}{d} n = 0$. Thus r is a zero divisor.

Now suppose that $r \in_n$ with $r \neq 0$ and that $\gcd(r, n) = 1$. If $rs = 0$ in *n* then in $\mathbb{Z}$ we have $rs = nk$ for $k \in$. Since r and *n* are relatively prime, *n* divides s and $s = 0$ in *n*. Thus r is not a zero divisor.

□

Corollary 88. *The ring $\mathbb{Z}_p$ with p a prime has no zero divisors.*

Proof. For p a prime number all of the elements $1, 2, 3, \cdots, p-1$ are relatively prime to p.

□

Definition:

Any commutative ring with unity, $1 \neq 0$, with no zero divisors is called an *integral domain*.

□

Example:

The following are examples of integral domains:

- By corollary ? the ring $\mathbb{Z}_p$ is an integral domain.
- The ring of integers $\mathbb{Z}$ is an integral domain. For $r, s \in \mathbb{Z}$, if $rs = 0$ then $r = 0$ or $s = 0$.

□

Theorem 89. *Every field is an integral domain.*

Proof. Let F be a field and let $a, b \in F$ with $a \neq 0$ and $ab = 0$. Since a^{-1} exists we multiply:

$$a^{-1}ab = a^{-1}0 \Rightarrow b = 0.$$

Therefore F has no zero divisors. Since F is a field it is a commutative ring with unity, and so F is an integral domain.

□

Of course not every integral domain is a field. The ring of integers $\mathbb{Z}$ is one such example. For another example of a class of integral domains we have the following theorem.

Theorem 90. *If R is an integral domain then the polynomial ring $R[x]$ is also an integral domain.*

Proof. Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n$ be polynomials over R with $f \neq 0$. Suppose that $fg = 0$. We consider the coefficients of f and g. Thus $a_mb_n = 0$. Since $a_m \neq 0$ and since R is an integral domain, $b_n = 0$. Similarly if $0 = a_mb_{n-1} + a_{m-1}b_n \Rightarrow 0 = a_mb_{n-1}$ and so since R is an integral domain $b_{n-1} = 0$. We continue in this fashion and see that $g = 0$. Therefore $R[x]$ is an integral domain.

□

Example:

The direct product of two integral domains is not an integral domain. Let R and S be integral domains and $1 \in R$ and $1 \in S$ unity. In the direct product $R \times S$ we have nonzero elements $(1, 0)$ and $(0, 1)$, whereas $(1, 0) \cdot (0, 1) = (0, 0)$.

□

We see that in the ring $\mathbb{Z}_n$ for any element $r \in \mathbb{Z}_n$ we have $nr = 0$. We might ask if this sort of property exists for other rings.

Definition: Let R be a ring. The *characteristic of a ring* is the least positive integer n such that $nr = 0$ for all $r \in R$. If no such integer exists, then the characteristic of the ring R is 0.

□

Example:

- The rings $\mathbb{R}, \mathbb{C}, \mathbb{Q}, \mathbb{Z}$ are all of characteristic 0.
- The ring $\mathbb{Z}_n$ has characteristic n

□

4.2.1 Exercises

1. Find the characteristic of the ring $\mathbb{Z}_5 \times \mathbb{Z}_4$.
2. Find the characteristic of the ring $\mathbb{Z}\times\mathbb{Z}$.
3. Let *R* be an integral domain. Prove that the characteristic of *R* is either 0 or a prime number *p*.
4. Describe the zero divisors of $M_2(\mathbb{Z})$ (2 × 2 matrices with integer entries).
5. Prove that every finite integral domain is a field.

4.3 Ideals

For further study of rings we recall the twists and turns from our study of groups. We have already seen that there are many analogs between groups and rings. Just as there are subgroups of a group, a subset of a group *G* that is a group under the same binary operation, we could study subrings of a ring. A subring is a subset of a ring that is a ring itself under the same binary operations. Despite the existence of subring structures, this is not a very important subject in the study of rings. A topic that we can get some mileage out of shares a connection with normal subgroups. The importance of these types of subgroups is that we can form quotient groups. There is an analog of these structures in the setting of rings.

4.3.1 Basic Properties of Ideals

Definition: Let *R* be a ring and *I* be a subgroup under the operation of addition. If for all $a, b \in R$ and $n \in I$, $an \in I$ and $nb \in I$ we say that *I* is an *ideal* of *R*.

□

Example:

For any ring *R*, clearly the ring itself is an ideal. Another ideal that every ring *R* possess is the group $\{0\}$ of the additive identity alone.

□

Theorem 91. If *A* is an ideal that contains a unit the $A = R$.

Proof. By definition $A \subseteq R$. Let $u \in A$ be a unit. There exists a multiplicative inverse $u^{-1} \in R$. For any $r \in R$, the element $r = r(u^{-1}u) = (ru^{-1})u \in A$. Thus $R \subseteq A$ and $R = A$.

□

Definition: Let *A* be an ideal of *R*. If $A \neq R$ and $A \neq \{0\}$ then *A* is *proper*.

□

Definition: For $a \in R$ the ideal $(a) = \{x \mid x = ra, r \in R\}$ is called a *principal ideal.*

□

Example:

In the ring of integers $\mathbb{Z}$, the principal ideal $(2) = \{\cdots, -6, -4, -2, 0, 2, 4, 6 \cdots\}$, the ideal $(3) = \cdots, -6, -3, 0, 3, 6, \cdots\}$. Every ideal in this ring is a principal ideal.

□

Definition:

If every proper ideal of the integral domain R is principal then R is called a *principal ideal domain*.

□

Example: Let F be a field. The integral domain $F[x]$ is a principal ideal domain.

4.3.2 Prime and Maximal Ideals

Definition: An ideal P is *prime* if for every $ab \in P$, $a \in P$ or $b \in P$.

□

Definition: The ideal M in the ring R is *maximal* if $M \neq R$ and for every ideal $M \subseteq N$, $N = M$ or $N = R$.

□

Theorem 92. *If M is a maximal ideal then it is also a prime ideal.*

Proof. Suppose $ab \in M$ and $a \notin M$. The ideal $(a) + M$ must be equal to *R* itself, due to *M* being maximal. So $1 = ra + m$ for some $r \in R$ and $m \in M$. Multiply each side of this equation by b and we have $b = rab + mb \in M$. Therefore *M* is prime.

□

Theorem 93. *If $\phi : R \to S$ is a homomorphism of rings and B is an ideal of S then $A = \phi^{-1}B$ is an ideal of R. If B is prime, then so is A. If ϕ is onto and B is maximal then A is maximal.*

Proof. The proof is left as an exercise.

□

4.3.3 Quotient Rings

Let *I* be an ideal of the ring *R*. The ideal *I* is a normal subgroup of *R* under addition. We define an equivalence relation on *R* by $a \equiv b \mathrm{mod} I$ if and only if $a - b \in I$. The equivalence class of $r \in R$ is $r + I = \{x \in R \mid x - r \in I\}$.

Definition:

Let *R* be a ring and *I* an ideal of *R*. The ring of equivalence classes under the above equivalence relation is called a *quotient ring* and is denoted R/I. Addition is defined as $(r + I) + (s + I) = (r + s) + I$ and multiplication as $(r + I) \cdot (s + I) = (rs) + I$.

□

Theorem 94. *The quotient ring R/P is an integral domain if and only if P is a prime ideal in R.*

Proof. If R/I is an integral domain then the ideal (0) is prime in R/I. The mapping $\phi : R \to R/I$ is an onto homomorphism of rings. So by theorem 93 $\phi^{-1}(0) = I$ is a prime ideal.

Suppose *P* is a prime ideal. If $(a + P)(b + P) = 0 + P$ in R/P then $ab \in P$. Thus $a \in P$ or $b \in P$. Thus $a + P = 0 + P$ or $b + P = 0 + P$, and R/P is an integral domain.

□

Theorem 95. *The quotient ring R/M is a field if and only if M is a maximal ideal in R.*

Proof. If R/M is a field then (0) is maximal in R/M. The mapping $\phi : R \to R/M$ is an onto homomorphism of rings. So by theorem 93 $\phi^{-1}(0) = I$ is a maximal ideal.

If M is maximal then $a + M = 0$ in R/M if and only if $a \in M$. If $a \notin M$ then $M + (a) = R$. Thus $1 = ab + c$ where $b \in R, c \in M$. Thus $(b + M)(a + M) = ab + M = (1 - c) + M = 1 + M$, which is the identity of R/M. Therefore the multiplicative inverse of every nonzero $a + M$ is given by $b + M$ and R/M is a field.

4.3.4 Exercises

1. Find all ideals of the ring $\mathbb{Z}_{18}$ and determine each of the quotient rings $\mathbb{Z}_{18}/I$.
2. Show that the only ideals in a field F are F and (0).
3. Find the prime and maximal ideals of the ring $\mathbb{Z}_{24}$
4. Find the prime and maximal ideals of the ring $\mathbb{Z}_3 \times \mathbb{Z}_3$
5. Prove that every prime ideal in a finite commutative ring with unity is also maximal.
6. Given that $\phi : R \to S$ is a homomorphism of rings
   a) Prove that if B is an ideal of S then $A = \phi^{-1} B$ is an ideal of R.
   b) Prove that if B is prime, then so is A.
   c) Prove that if ϕ is onto and B is maximal then is maximal.

5 Bibliography

[1] B. Baumslag and B. Chandler, "Group Theory," McGraw Hill, Inc., New York, 1968.

[2] A. Clark, "Elements of Abstract Algebra," Dover Publications, Inc., New York, 1984.

[3] D. Dummit and R. Foote, "Abstract Algebra," John Wiley & Sons, Inc., Hoboken, NJ, 2004.

[4] J. Fraleigh, "A First Course in Abstract Algebra, 7th Edition" Addision Wesley, New York, 2003.

[5] T. Hungerford, "Algebra," Springer-Verlag, Inc., New York, 1974.

[6] N. McCoy, "Rings and Ideals" Mathematical Association of America, Baltimore, MD, 1948.