An Operator Manual:

An introduction of abstract algebra for contemporary contexts

James B. Wilson

March 4, 2016

Copyright ©2011. James B. Wilson
Department of Mathematics
Colorado State University
101 Weber Building
Fort Collins, CO 80523

jwilson@math.colostate.edu

Written with LaTeX 2_ε.

Please Recycle when finished.

Contents

Counting to the Octonions

Our first experience in mathematics is to count. So the numbers $1, 2, 3, \ldots$ have the right to be called *natural*. Addition follows soon after as a means to explain how a count to five can always we be accomplished by counting to three and then counting two more. Problem solving begins with the introduction of variables. When is there an x for which $3+x = 5$? Subtraction and negative numbers make an appearance both because they help us understand concepts such as debt and also for the simple usefulness in solving equations: $x + 2 = 5$ so also $x = 5 - 2$. By this point we have developed the integers. Indeed, multiplication seems obvious as a shortcut to adding up the same number several times. Indeed, we go often say "three *times* five". As with addition, the invention of multiplication introduces problems where an inverse process would be helpful, which we do by inventing division. Thus $3x = 15$ is solved by dividing $x = 15 \div 3$. We also find direct value in understanding the implied new numbers $1/3$, etc. so it is not too difficult to accept fractions as numbers.

Powers mimic the path from addition to multiplication in that powers are used as a short-hand for multiplying the same number many times, e.g. $3^4 = 3 \cdot 3 \cdot 3 \cdot 3$. So defined, exponents make sense only for natural numbers. However, we soon discover some sensible patterns including:

$$a^{n+m} = a^n a^m \tag{1}$$

$$a^{nm} = (a^n)^m. \tag{2}$$

This suggests the logical meaning for a^{-1} should be whatever number has the property that $a^n a^{-1} = a^{n-1}$. The solution is to set $a^{-1} = 1/a$ and along with that $a^0 = 1$. Now we have exponent with integer powers. Next up we consider fractional powers. Here $a = a^1 = a^{m/m} = (a^{1/m})^m$ is the guide. We notice that the meaning of $a^{1/m}$ is a solution to $x^m = a$. Therefore, to introduce fractional powers is the same as introducing solutions to $x^m = a$. Unfortunately, these solutions are not always within our current number rational number system.

Theorem 0.0.1. *There is no rational number x such that $x^2 = 2$.*

Proof. Suppose the claim is wrong. It would mean there are integers a and b such that $\left(\frac{a}{b}\right)^2 = 2$. We know that fraction can be reduced until a and b have no common multiples, save ± 1, and so we assume that is the case. Now we know $2b^2 = a^2$. As 2 is prime, 2 divides a^2 (i.e. a^2 is even). Thus 2 must also divide a (i.e. a is even). Since $a = 2k$ for some integer k it follows that $2b^2 = a^2 = 4k$ so that $b^2 = 2k$. This forces b to be even. Now both a and b have 2 as a factor. We know this is not the case. This contradiction implies our lone assumption is not correct, that is, 2 is not the square of a rational number. □

We are capable of creating new numbers which extend the rationals to include a solution to the equation $x^2 = 2$, and we typically call this number $\sqrt{2}$, though it is not unique as $-\sqrt{2}$ is also a solution. However, that process will

ultimately waste our time as with a simple modification we encounter an infinite diversity of equations without rational solutions.

Lemma 0.0.2 (Euclid's Lemma)**.** *If p is a prime and a and b are integers such that ab is multiple of p, then either a is a multiple of p or b is a multiple of p.*

Using Euclid's Lemma we may prove the following as well.

Corollary 0.0.3. *For every positive integer m and every prime p, the equation $x^m = p$ has no rational solutions.*

Therefore to continue our construction of necessary numbers we need a different inspiration. Through the Pythagorean Theorem we recognize that a right triangle with two sides of length 1 has a hypoteneus of length $\sqrt{2}$ and so the existence of $\sqrt{2}$ is unavoidable in geometry. So we should think about numbers not only as consequences of solving algebraic equations but also as lengths in geometry and so discover a method to model all numbers.

So we pause and consider what length should require. First, every natural number n already represents a length so $|n| = n$ for natural numbers. If we think of negatives solely as imposing direction then $|-n| = n$ and $|0| = 0$. If we double an integer it doubles in length, and so on, so in general $|n \cdot m| = |n| \cdot |m|$. Finally, if the lengths of triangles are to be considered then the hypoteneus of a triangle has a length shorter than the sum of the other two sides and that explains why $|m + n| \leq |n| + |m|$. So we take those three observations about length to be the assumptions we require of all numbers.

The existence of length enables us to consider approximations of the numbers we are after using ones we already have. Our experience with calculators makes this easy: we can zoom into the graph of $y = x^2 - 2$ as much as we wish and see the ever improved estimate for $\sqrt{2}$. This leads to:

$$\sqrt{2} = 1.4142\cdots = 10^0 + 4 \cdot 10^{-1} + 1 \cdot 10^{-2} + 4 \cdot 10^{-3} + 2 \cdot 10^{-4} + \cdots$$

and this sequence of digits never settles into a repeating pattern (otherwise it would be rational). Eventually all the roots we considered above are described in this way.

It takes a great deal of work to show that every root of the form $x^m = p$, for p a prime, occurs as a decimal number. This is commonly solved using Newton's method in Calculus. Regardless of the precise method used all such proofs depend in some way on the results about continuous functions, for example, the Intermediate Value Theorem. Because of that it is often claimed that Algebra depends on Calculus. While this perspective has some merit there is an error. When we create decimal numbers we are creating many many numbers which are not strictly necessary for algebra (though they are certainly crucial for geometry). These are known as *transcendental* numbers and they include numbers such as π and e. If we restrict attention to just the 'algebraic' numbers we no longer need calculus to make progress. However, that would then require we go back the tedious method of extending the numbers with each new root one by one. So this is a Pyrrhic victory if it is a victory at all.

Finally, it has not escaped our notice that $x^2 + 1 = 0$ has no solution as a decimal number. For if x is a decimal number then $x^2 \geq 0$ and so $x^2 + 1 > 0$. For this equation to have a solution thus requires yet more numbers, numbers that include $i = \sqrt{-1}$, and as seen in the quadratic formula:

$$ax^2 + bx + c = 0 \Rightarrow x = \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a} = \begin{cases} \frac{-b}{2a} \pm \frac{\sqrt{b^2-4ac}}{2a} & b^2 \geq 4ac \\ \frac{-b}{2a} \pm \frac{\sqrt{4ac-b^2}}{2a} i & b^2 < 4ac. \end{cases}$$

At this point we have created the *complex numbers* $\mathbb{C}$, the real numbers $\mathbb{R}$, the rational numbers $\mathbb{Q}$, the integers $\mathbb{Z}$, and each of these is demanded by our need to count with the natural numbers $\mathbb{N}$.

Theorem 0.0.4 (Gauss' Fundamental Theorem of Algebra)**.** *For all natural numbers n and all complex numbers $\alpha_0, \alpha_1, \ldots, \alpha_n$, there are n complex numbers $z_1, \ldots, z_n$ unique to the α_i such that*

$$\alpha_n x^n + \alpha_{n-1} x^{n-1} + \cdots + \alpha_1 x^1 + \alpha_0 = \alpha_n (x - z_1)(x - z_2) \cdots (x - z_n).$$

In particular, each z_i is root of the polynomial.

Gauss' theorem is called the Fundamental Theorem of Algebra because it has a significant and definitive conclusion to the methods of classical algebra of polynomials. Yet, algebra at the time of Gauss was about to change completely owing in large part to his own work and also to the work of two young an tragic figures: Abel and Galois.

It was already evident to Lagrange that complex numbers would be sufficient to explain the roots of polynomials up to degree 4. However, Lagrange had discovered that his method was not capable of producing the roots of polynomials of degree 5. The tools of his day were largely Calculus – rather Analysis as we now describe it. Finally, using Analysis Abel finally proved that Lagrange and others were of the right intuition.

Theorem 0.0.5 (Abel's Unsolvability of the Quintic)**.** *There is no "radical formula", (a formula involving only addition, multiplication, and the taking of roots) which describes the roots of a polynomial of degree five in terms of its complex coefficients.*

So although Gauss had proved the complex numbers have all the roots, Abel's result explained that it was impossible to describe what the roots truly are in terms of formulas such as the quadratic formula does for quadratic polynomials. He had sent his proof the Cauchy – the inventor of Analysis (modern calculus) but that failed to secure Abel the notice he deserved and he died shortly after in poverty. However impressive, Abel's proof is no longer considered. In its place is a nearly contemporary solution by Galois. The reason Galois' proof survives is that it gave a convincing and versatile process by which to understand which polynomials have no solution by radicals.

Theorem 0.0.6 (Galois' Unsolvability Criteria)**.** *A polynomial has a radical formula for its roots if, and only if, there is a sequence of extensions of numbers having at each stage all the roots of an polynomial that divides the original.*

We later come to name these extensions *normal extensions* and they include examples like adjoining $\sqrt{2}$ to the rationals, since that also adjoins $-\sqrt{2}$ so it factors $x^2 - 2$ completely. However, adjoint $\sqrt[3]{2}$ is not such an extension because

$$x^3 - 2 = (x - \sqrt[3]{2})(x^2 + ax + b)$$

does not completely factor with those numbers.

Having learned to accept and create new numbers with such range we should ask, why stop here?

Algebra and geometry are amongst the earliest pillars of mathematics appearing long before Calculus, Set Theory, Graph Theory, Computer Science, etc. This is perhaps because they address two fundamental problems. Algebra is used to solve for unknown quantities whereas geometry is used to measure shapes. To do such mathematics demands numbers with beneficial properties. The standout properties include that equations such as $3 + x = 5$ and $3x = 15$ have a single solution. So we expect numbers $\alpha, \beta, \ldots$ will satisfy:

- there is a unique x such that $\alpha + x = \beta$, and
- if $\alpha \neq 0$ then there is a unique x such that $\alpha x = \beta$.

In making numbers that get after these two properties we might need to include some numbers we were not expecting, e.g.: $5 + x = 3$ requiring us to create -2 (which is not a number used to count or measure). Hence, to make sense alongside of geometry we need each number α have a magnitude $|\alpha|$ which is once more a number we recognize as a measurement, for example, $|-2|$ becomes 2 which is something we can count and measure. We will also need to compare one magnitude with another to get concepts of length and distance. So we need exactly one number of no length.

- $|\alpha| = 0$ if, and only if, $\alpha = 0$.

To tie in with algebra, we should be able to compare the length of addition and multiplication as well. Indeed, formulas for area explain how this might be done. For instance area of a rectangle is length times width. The formulas might also be modified by constants, e.g. the area of a triangle is only $1/2$ length times height, and the area of a circle is the radius times the radius times the constant π. The essential property is captured as follows: the magnitude of a product is a product of magnitudes, or in symbols:

- $|\alpha\beta| = |\alpha||\beta|$.

One final demand of magnitude is we can compare the length of a sum of numbers to the their individual lengths. For instance, in a triangle the sum of two sides cannot be longer than the third. So we might consider the triangle inequality:

- $|\alpha + \beta| \leq |\alpha| + |\beta|$.

There are few technical glitches in what we have written so far. For example we used 0 as a number without explaining what it is (whatever it is at least it is the unique solution to $\alpha + x = \alpha$ and also the unique number of magnitude 0). Such issues can all be addressed but the central issue comes down this:

What are the numbers that work as just described?

The answer may surprise us. For even though we are comfortable and familiar with decimal numbers and possibly aware of imaginary numbers, that is not the answer. In fact the answer is 8 times larger than decimal numbers (in that they would take an 8 dimensional universe to draw). They are called *Octonions*. To reach them we need to start from the beginning.

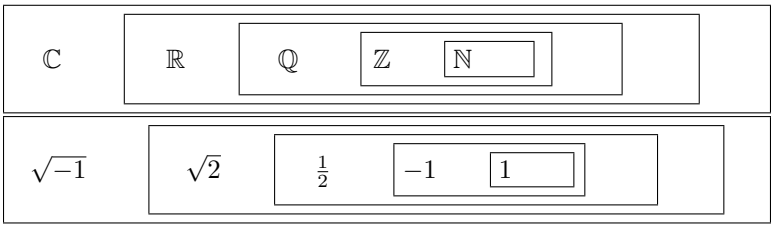


Figure 1: On the top we see how the classic sets of real numbers are nested; on the bottom we see an element in each greater set that shows why the sets are not all the same.

Chapter 1

Models of algebra

Motivation

In this chapter we will introduce many of the fundamental examples of operations on sets. Operations include familiar ideas such as adding number, composing functions, and subtracting matrices. We will also see they explain other concepts such as the numbers 0 and 1 as well as negatives and inverses. There are even operators that take three number or matrices and return just one.

There are too many operations to handle each in detail, but there is tremendous value in comparing one family of operations against another. Thus our list is rather longer than might be found in traditional settings. Our taxonomy will include semigroups, monoids, groups, loops, rings, fields, and semifields. This list contains superstars and runts and we accordingly give more attention to the stars. To lay the foundation that allows us to study each of these efficiently and uniformly we review some elements of Set Theory introduce the concepts of Model Theory and Universal Algebra as they pertains to defining varieties of algebraic operations.

1.1 Sets and classes

Further Reading: Jacobson §0.1-0.2

To start, a *set* is a collection of items, objects, substance, etc. For instance, "the set of all cars" or "the set of all integers". We speak of sets as containing other things, such as "the set of all cars contains the Delorean" or "the integers $\mathbb{Z}$ contain 0" and we write $0 \in \mathbb{Z}$. Some sets will not contain anything, we call those *empty*. We will see later that the empty sets, much like the Delorean, is a rare object indeed (more in Exercise 2.1.1). We also have a related concept of a *class*, which to us will mean a property, for instance, "the class of cars that get 40 miles per gallon on the highway."

Many classes actually describe sets, for instance there is a set of all cars that get 40 miles per gallon on the highway. However, there will be cases were a class does not describe a set. As we might expect, the difference occurs once we become more specific about what we want sets to mean. One of the specifics is whether or not a set can list itself as a member. This makes sense in some contexts, for example, a book can contain other books – for example *The Complete Works of William Shakespeare*. Yet, it is not sensible to expect a book to contain *itself* because if it did attempt this feat it would be an unending process. However, some would argue that self-containment is not so impossible and point to examples like a house of mirrors which reflects the same image over and over forever getting smaller and smaller.

Theorem 1.1.1 (Russell's Paradox)**.** *There is no set of all sets that do not contain themselves. In particular, there is a class that has no associated set.*

Proof. Suppose S is the set of all sets that do not contain themselves. We have made no rule about whether or not a set can contain itself and we can either have $S \in S$ or $S \notin S$. But only one of these can be true and so we must decide.

If $S \in S$ then S is a set that does not contain itself (by the definition of what it means to be in S). However, this fights the assumption that $S \in S$. So we are left to conclude that $S \notin S$. But in that case S is set that does not contain itself and so $S \in S$. That contradicts the conclusion that $S \notin S$.

The result is that our very *first* assumption must be false. Hence we *cannot* suppose that there is a set of all sets that do not contain themselves (as we did in our first line of our proof).

For the final claim, notice "all sets that do not contain themselves" is a property and so there is a class here which is not associated to a set. □

Whatever one's personal perspective, Russell's Paradox makes it clear that there are problems with sets containing themselves and these problems are not simply that we enter an infinite feedback loop. Rather logic itself does not allow us to do this unless we are exceptionally careful with our definitions. The simplest solution is the one taken by most mathematicians: we simply do not allow sets to contain themselves. Classes on other hand do not actually contain anything, we just use that vocabulary with classes because it is convenient.

We will often attempt to treat sets and classes as the same because the arguments we make with sets and classes often have identical logical underpinnings. For example, we can talk of the *union* of sets A and B, denoted $A \cup B$, which means the collection of elements that come from A *or* B. Of course the word *or* can also be used to combine two properties: the class of all pigs or the class of all donkeys. Hence, it makes sense to think of this as the union of the class of all pigs and the class of all donkeys. This way, if a class does give rise to a set then we have no need to argue with new vocabulary. Notice intersections rely solely on the word *and* and so we can intersect both sets and classes without harm. Finally, we can speak of "an element of a class" and here we simply mean the property of the class is satisfied by the candidate element. Because of the overlap in vocabulary and meaning we will find that many definitions for sets also make sense for classes, for instance, the notion of a function between sets leads to a notion of functions between classes as well.

The words 'collection' and 'property' are left undefined purposefully but they can be given a more rigorous treatment. Tame introductions include Tarski's *Introduction to Logic*, Dover Publications Inc., New York, and Halmos' *Naive Set Theory*, Springer, New York. Even so, those works will leave certain terms undefined as well. We should not think of this as a failure in mathematics. To the contrary, mathematics is a conversation about ideas and so it must be carried out in our languages, in our case English. What we first learn about language only permits us to explore concepts in the center of mathematics, such as counting. But as we develop more language skill we can reason further. This reasoning is in *all* directions. That is, we might take our counting and turn it into fractions and decimals, or we might begin to think how numbers are artificial place holders for sets of the same size. In this way, mathematics is not truly 'axiomatic' in the sense that we make a specific 'best possible' list of assumptions and consider only the implications from those assumptions. Indeed, mathematics is also continually works away from axioms by replacing them with ever more subtle and complex assumptions which help expose opportunities for new fields of reasoning.

1.2 Relations and Functions

Further Reading: Jacobson §0.3

Sets and classes allow us to formalize many concepts once known intuitively, for example unions are a formal treatment of combining to properties. The concept of a relation allows us to consider certain sets in context to bigger sets. For example, we might wish to consider how car color relates to resale value. This could be captured by a table in which one column lists the possible colors of a car and the second columns list the average resale value for a car of that color. If we strip away the meaning of colors and dollars we find we have made a list of ordered pairs $(\text{color}, \text{value})$ and the entire table is therefore a subset of all possible ordered pairs (a, b) where a is color and b is number. This general idea may seem to have lost the point of the original problem: we have removed the meaning from the set. However, by doing so we can stand back and see in an easier way that this set is similar to other sets. For example, this set might be

compared with a table relating colors in TV comercials to average number of viewers who respond positively to the advertisment. By releasing the semantics (i.e. the meaning) from the two sets we can make beneficial connections between the tables.

A *relation* on non-empty sets (or classes) S and T is a subset (or subclass) R of the set (or class) $S \times T$ of ordered pairs (s,t), for $s \in S$ and $t \in T$. If $(s,t) \in R$ we will write sRt. This 'infix' notation attempts to capture the familiar relations we use elsewhere. For example, if $S = T = \mathbb{Z}$ then we can write $a \leq b$ for $a, b \in \mathbb{Z}$. This common relation can be captured by our somewhat awkward definition above by letting

$$R = \{(a,b) \in \mathbb{Z} \times \mathbb{Z} : b - a \text{ is a natural number } \}.$$

So when we write aRb we mean that $b - a$ is a natural number and so in our more familiar notation $0 \leq b - a$. Hence, aRb is the same as writing $a \leq b$. What we manage to do by using R instead of $\leq$ is that we have not required any external knowledge about the ordering of numbers. Instead, we just require a knowledge of subtraction and of natural numbers. This makes a logical progression possible. First we seek to create natural numbers, then we create integers along with subtraction. The implied ordering then comes out of that process. We shall expand on that construction later.

Remark 1.2.1. As this is our first formal definition I need to emphasize an unspoken protocol in mathematics. A definition is simply assigning a name or notation to a series of objects or properties (sets or classes). This assignments is always meant to be in both directions. For example, if E is a relation on A and B A, then R must be subset of $A \times B$. Likewise, if E' is a subset of $A \times B$ then we call E' a relation. Despite obvious logical and grammatical flaws, we often capture this assumption by the phrase:

All definitions are "if, and only if,".

It is also crucial to understand that no definition is ever in need of a proof.

One of the most powerful relations is that of a function. Indeed, functions are perhaps the best known mathematical object after numbers and so we hope that our following technical description will not supplant the intuition gained by years of experience with functions. However, as we move our reasonnig increasingly further away from real numbers and elementary functions that we can graph, it becomes necessary to declare the formal expectations about functions.

The most familiar definition of a function is a relation from a set A to a set B for which every $a \in A$ is related to one and only one $b \in B$. We want also the intermediate concept of a partial function.

Definition 1.2.2. A *partial function* $f : A \to B$ between sets (or classes) A and B is a relation $R \subseteq A \times B$ such that:

if for some $a \in A$ and $b, b' \in B$, aRb and aRb', then $b = b'$.

We say that R is *well-defined* when that property holds and we write $f(a) = b$ since no ambiguity can occur about which element of B to associate to a.

- The *domain*, denoted $\operatorname{dom} f$, of a partial function is $\{a : (a,b) \in R\}$.
- The *image*, denoted $\operatorname{im} f$, of a partial function is $\{b : (a,b) \in R\}$.
- The *codomain*, denoted $\operatorname{codom} f$, of a partial function is B.

Finally, a *function* is a partial function whose domain is A.

1.3 Operations on Sets

Having described sets and functions we now consider a very special setting where we have functions between a common fixed set. This is quite common in mathematics upto Calculus because we often consider only the functions of the form $f(x) = 3x + 9x^2$ or $g(x) = \sin x$ and these always begin and end with the set $\mathbb{R}$. However, we are even more interested in functions such as $f(x, y) = x+y$ or $f(m, x, b) = mx + b$.

Definition 1.3.1. An *operation* on a set S is a partial function from an arbitrary number of copies of S into S.

- An operation $* : S \to S$ is called *unariy* and is often written in postfix or exponential notation x^*.
- An operation $\cdot : S \times S \to S$ is called *binary* and usually written in infix notation $x \cdot y$.
- An operation $S^n \to S$ is an n-ary operation and in generaly has no standard notation, but we will denote it by $\{s_1, \ldots, s_n\}$.

We are most aware of unary and binary operations. For example, on the real number $\mathbb{R}$ we have the unary operation $x \mapsto \frac{1}{x}$, or rather x^{-1}. This operation is not defined at 0 and so it is only a partial function on $\mathbb{R}$. Other familiar unary operations on $\mathbb{R}$ include $x \mapsto x^n$ for arbitrary n. Of course, if n is a faction such as $1/2$ we may encounter values without a square-root and so again these are only partial functions.

Several common binary operations in $\mathbb{R}$ include addition, mutliplication, exponentiation, subtraction, division, etc. Notice addition, subtraction, and mutliplication are functions but division and exponentiation are only partial functions. For example, we do not encounter division by 0 in the real numbers and some exponents are either complex or somewhat questionable to describe (for example $(-1)^{\sqrt{2}}$). We shall see in a moment that ternary operations also make sense for real numbers and are an incredibly efficient (dare we call them clever) tool in proofs.

It is customary to introduce operations as $\langle S, * \rangle$ where $*$ is the operation and following that write only S. For example, we write $\langle \mathbb{R}, + \rangle$ the first time we consider addition on $\mathbb{R}$ but afterward we abbreviate this to $\mathbb{R}$. In fact, it is quite common to see authors announce that convention by writing the logically impossible sentence: $\mathbb{R} = \langle \mathbb{R}, + \rangle$. Of course, those authors understand that $\mathbb{R}$ cannot reference itself in that way but they are simply explaining to the reader that they are temporarily dismissing all other known properties of $\mathbb{R}$ to focus exclusively on the set with its operation of addition.

Finally, for many the study of operations was constrained to mathematics. However, an explosion in interesting unary and binary operators came out of the development of robust computer programming languages such as `C++` and `Java`, e.g. `++i`, `k--`, `s<<2`, `s || t`, and `s ? a : b`.[1] For that reason it is increasingly useful to study operations without assumptions on their properties. Hence, we postpone a tour through the usual properties (e.g. associativity and commutativity) until after we discover the essential properties of congruence.

[1] These notations produce $i + 1$, $k - 1$, shifting the digits of a binary number to the left by 2 (same as multiplying by 4 but done very efficiently), taking bitwise or of the bits in s and t, and the final ternary operation outputs a if $s \neq 0$ and otherwise b. That can be viewed as an advance variant of Kronecher's δ function in mathematics.

1.4 Models

In this section we introduce the meaning of *axioms*, *theorems*, and *proofs* in the context of abstract algebra. It is perhaps in algebra that for the first time we encounter proofs of properties that actually have no meaning on their own but must later be associated to a specific context to gain meaning. For example, in Calculus we prove the Mean Value Theorem and that gives an immediate description of how the tangents of a curve relate to the overall change of the curve. However, in abstract algebra we make claims such as,

> "If we can divide by zero then every number equals 0."

Of course that sentence is immediately pointless because we never allow division by zero. Yet, what this abstract scenario does for us is provide a sand-box to try out why we do not wish to permit division by zero. We find that it has little to do with how we might slice a pizza (after all, no one can slice a pizza into $1/\sqrt{\pi}$ parts even though every calculator can divide by $\sqrt{\pi}$ without effort). Instead, we find that the consequence is severe: all numbers must be equal to zero for division by zero to be possible. Thus, this sentence is vitally important as it indicates the reason we never use number systems which would permit division by zero.

1.4.1 Axioms and Theorems

Stepping back for a moment, we are not invested in understanding the philosophical implications of logic as a whole so we ignore sentences such as "The cow jumped over the moon." Instead sentences (or statements) here will mean only mathematical content such as

> "The variable x is equal to three." (abbreviated as $x = 3$).

In principle the sentences we care about can be assigned a value of true or false but perhaps require some context to make the assignment. For instance, $x = 3$, is neither true nor false until we interpret the meaning of x. So we might say $x = 5 - 2$ *satisfies* or *models* the sentence $x = 2$ because the interpretation of x is now made clear and indeed in that interpretation the sentence becomes true. On the other-hand, $x = 5 + 2$ does not satisfy (is not a model for) $x = 3$.

There are some subtle points in this formality. E.g.: the reader sees $\star$ and ε instead of $+$ and 0, or $\cdot$ and 1. This means that the application of a theorem which follows from a sentence in abstract algebra is not meaningful until we substitute in not only variables (which is by now quite common in our level of mathematical experience) but we must also substitute in for the *operations* and any constants we deem essential, such as replacing the identity ε with 0 or 1. However, we do not substitute in a value for s. That distinction can become very difficult to perceive as we get more involved sentences. Thus, we need to be more precise in describing our variables.

First consider the definition of $n!$. This is usually described as $n! = \prod_{i=1}^{n} i$. Here the role that n plays as a variable is different form the role for i. The variable n must be given to us at a later time, for example 5! replaces n with 5. However, the person interested in 5! will not provide us with a value for i, rather the sentence informs us what values i should take once we know a value for n. This is because the variable i is *bound* to the sentence defining $n!$ where the letter n is *unbound*. If you are familiar with programming you may have encountered a difference between so called *global* and *local* variables. This is the same situation. Unbound variables (global variables) are not explained by the content of the sentence and so they must be interpreted before using the

sentence. On the other hand bound variables (local variables) are there to do work for the meaning of the sentence and so bound variables are specified with quantifiers such as for all, $\forall$, and there exits, $\exists$, or for all i between 1 and n, etc.

So let us begin by abstracting the usual idea of zero and one. There is only one number α such that $x + \alpha = x$ (i.e. $\alpha = 0$) and similarly there is only one number β such that $\beta x = x$ (obviously $\beta = 1$). We can prove these two as independent facts but they are naturally explained by one common principle. This is because the values 0 and 1 in our usual numbers systems are *models* of the next sentence.

Identity Axiom

For a binary operation $\star$ on a set S there is $\varepsilon \in S$ where for all $s \in S$,

$$s \star \varepsilon = s = \varepsilon \star s.$$

Such an element ε is called an *identity* for $\langle S, \star \rangle$.

To complete our illustration we prove the ε above is unique.

Theorem 1.4.1. *If $\langle S, \star \rangle$ satisfies the identity axiom then the implied identity is unique to $\langle S, \star \rangle$. Hence we now speak of* the *identity of $\langle S, \star \rangle$.*

Proof. Let $\varepsilon, \varepsilon' \in S$ be elements of S where for every $s \in S$, $s \star \varepsilon = s = \varepsilon \star s$ and $s \star \varepsilon' = s = \varepsilon' \star s$. It follows that

$$\varepsilon = \varepsilon \star \varepsilon' = \varepsilon'.$$

□

Now in order to write an axiom that properly separates the unbound and bound variables we introduce terminology from *Universal Algebra*.

Definition 1.4.2. A *signature* is a list of operations types that may be applied to describe the operations an a class of sets with operations. Such a set will be called a σ-algebra.[2]

Remark 1.4.3. Since a signature records only the types (e.g. binary, unary, etc.) of operations we wish to impose on a set it is entirely sufficient that a signature be a list of numbers, such as $[2, 2, 1, 0, 0]$ instead of $[+, \cdot, -, 0, 1]$. We prefer the intuition communicated by the symbols.

Example 1.4.4. (i) The signature of the natural numbers under addition would be $[+, 1]$ where $+$ is a binary operation and 1 is a "nullary" operations – i.e. a constant.[3] We distinguish 1 as it is necessary to define all the natural numbers, e.g. $1, 1 + 1, (1 + 1) + 1, \ldots$. Notice that 1 is not the identity for addition of natural numbers. This illustrates how a signature does not impose semantics (meaning) on the symbols.

(ii) The signature of the integers under addition and multiplication is $\sigma = [+, \cdot, -, 0, 1]$ where $+$ and $\cdot$ are binary, $-$ is unary (corresponding to taking negatives), and 0 and 1 are nullary. In principle, a statement concerning

[2] There is an even older use of the name σ-algebra that occurs in *Measure Theorey*: the construction of modern integrals. These two meanings are unrelated.

[3] Recall an n-ary operator on a set S is a function $S^n \to S$. So a nullary operator should mean a function $S^0 \to S$. There are two sensible interpretations of S^0. Either S^0 is the empty-set as it represents an empty product, or $S^0 = \{()\}$ the set with the empty coordinate. With $S^0 = \{()\}$ we have the size of S^0 equal to 1 similar to how we choose $2^0 = 1$, etc. It is in that sense that a nullary operation $1 : S^0 \to S$ corresponds to a constant, namely there is only one element in S which is in the image of the function.

integer addition or multiplication can be expressed using the symbols in sigma and then general logical notation such as variables, parentheses, conjunctions (and, or), not, and quantifiers $(\forall, \exists)$.[4] This signature also applies to the natural numbers even if $-$ and 0 are not used in that set. This is necessary because the integers cannot be described without natural numbers. In this situation the natural numbers treat $-$ and 0 as partial functions (i.e. the domains are only subsets of $\mathbb{N}^2$ and $\mathbb{N}^0$ respectively).

It has become convention to use a signature that reflects the implied semantics of the symbols so that the reader can borrow from intuition. So, the identity axiom can now be written precisely as follows:

Identity Axiom (redux)

S is a set with a binary operation $\cdot$ and a nullary operation 1 such that

$$\forall s \in S, \quad s \cdot 1 = s = 1 \cdot s.$$

The unbound variables, S, $\cdot$, and 1, are all specified at the start without any logical quantifiers whereas the bound variables are preceded by a logical quantifier, in this case $\forall$.

Remark 1.4.5. We caution the use of 1 instead of ε and $\cdot$ has pedagogical implications. For example, to say that the the integers model the identity axiom would ask us to write $\cdot = +$ and $1 = 0$. Though this is certainly legal, it reflects a disrespect to how we tend to learn. It is best to be verbose in such settings saying that $+$ replaces the abstract $\cdot$ and 0 plays the role of 1.

Definition 1.4.6. A *model* of a set Φ of sentences is an assignment of variables, operations, and constants that satisfy each sentence in Φ.

With the vocabulary of models in place we can finally describe a reliable scheme for theorems in algebra.

Definition 1.4.7. Let ϕ and Φ be sentences with a common signature σ.

(a) A sentence ϕ is a *consequence* of a set Φ of sentences if: for every model M of Φ, M is a model for ϕ. We often say ϕ is a *theorem* following from Φ.

(b) The *theory* for Φ is as the class of all consequences of Φ. The sentences in Φ are the *axioms* (also known as *postulates* or *laws*) of the theory.

Notice that our definition of theorems does not try to capture the differences we impose with names like *lemma*, *proposition*, *corollary*, etc.. Those labels should be considered as theorems in the sense of model theory. We simply impose such names to give a hierarchy of the significance of a result. The hierarchy is somewhat personal and there are many examples in mathematics of truly significant lemmas (e.g. Schur's Lemma) and totally pointless theorems (no examples necessary).

Also notice that a theory associated to a set Φ of axioms might also be the theory associated to a different set Γ of axioms. We will see such a situation in an exercise below. This gives us some flexibility. We might begin with axioms we found through some natural means but along the way find reason to use alternatives. The theory is unaltered.

It is critical to be clear about which assignment of variables, operations, and constants we intend when we speak of a model for a set of sentences. The following example illustrates how different the results can be.

[4] This demonstrates that we are constraining ourselves to *first order* logic.

Example 1.4.8. (a) The natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$ do not model the identity axiom under addition (because, for all $m \in \mathbb{N}$, $1 + m \neq 1$ and so no number $m \in \mathbb{N}$ satisfies the identity axiom).

(b) The natural numbers $\mathbb{N}$ are model for the identity axiom, specifically where multiplication is the operation and 1 is the identity element.

1.4.2 Common Axioms of Algebra

The following is an incomplete list of the axioms which appear in algebra. We organize the list according to the signature and axioms at the top are more common than ones at the bottom. In most situations we combine several axioms and multiple operations. However, these constitute different and essential properties.

Axioms with signature $\{\cdot, ^{-1}, 1\}$.

Associativity Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s, t, u \in A, \quad (s \cdot t) \cdot u = s \cdot (t \cdot u).$$

Commutativity Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s, t \in A, \quad s \cdot t = t \cdot s.$$

Identity Axiom

For a $\{\cdot, 1\}$-algebra A,

$$\forall s \in A, \quad s \cdot 1 = s = 1 \cdot s.$$

Inverse Axiom

For a $\{\cdot, ^{-1}, 1\}$-algebra A, A satisfies the identity axiom and

$$\forall s \in A, \quad s \cdot s^{-1} = 1 = s^{-1} \cdot s.$$

Substitutes for associativity:

Moufang Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s, t, u \in A, \quad (s \cdot t) \cdot (u \cdot s) = s((t \cdot u) \cdot s).$$

(Left) Alternative Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s, t \in A, \quad (s \cdot s) \cdot t = s \cdot (s \cdot t).$$

Jordan Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s, t, u \in A, \quad (s \cdot s) \cdot (t \cdot s) = ((s \cdot s) \cdot t) \cdot s.$$

Flexible Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s, t \in A, \quad (s \cdot t) \cdot s = s \cdot (t \cdot s).$$

Power Associative Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s \in A, \quad (s \cdot s) \cdot s = s \cdot (s \cdot s).$$

Substitutes for commutativity.

Idempotent Axiom

For a $\{\cdot\}$-algebra A,

$$\forall s \in A, \quad s \cdot s = s.$$

Axioms with signature $\{+, \cdot, -, 0, 1\}$.

Distributive Axiom

For a $\{+, \cdot\}$-algebra A,

$$\forall s, t, u \in A, \quad s \cdot (t + u) = (s \cdot t) + (s \cdot u), \quad (s + t) \cdot u = (s \cdot u) + (t \cdot u).$$

Jacobi Axiom

For a $\{+, \cdot\}$-algebra A,

$$\forall s, t, u \in A, \quad s \cdot (t \cdot u) = (s \cdot t) \cdot u + t \cdot (s \cdot u).$$

Skew-commutative Axiom

For a $\{+, \cdot, 0\}$-algebra A,

$$\forall s \in A, \quad s \cdot s = 0.$$

1.1 For each equation below, choose the fewest combination of axioms from the standard list which allow you to state and prove the equation.

(a) $0 \cdot x = 0 = x \cdot 0$.

(b) $(-1) \cdot x = -x$.

(c) $(-x)(-y) = xy$.

(d) There is only one number that can be a 0 and only one number that can be a 1.

(e) If $ab = 0$ then either $a = 0$ or $b = 0$, or both.

Showing all your steps from the axioms can be tedious and usually is not helpful. For instance, in the following example most of the steps are trivial yet they require a lot of work to do formally. We will not generally have to see all these steps.

Example 1.4.9. Solve for x in the following equation making sure to indicate at each step what axioms you use. Do one step at a time and do not skip steps.

$$4 + 3x = 2x - 1.$$

Solution:

$$\begin{aligned}
4 + 3x &= 2x - 1 = 2x + (-1); \\
3x + 4 &= 2x + (-1); && \text{Commutativity of addiontion (left side.)} \\
(-2x) + (3x + 4) &= (-2x) + (2x + (-1)); && \text{Equality is a congruence for addition.} \\
((-2x) + 3x) + 4 &= ((-2x) + 2x) + (-1); && \text{Associativity of addition.} \\
((-2 + 3)x) + 4 &= ((-2 + 2)x) + (-1); && \text{Distributive Law.} \\
((-2 + (2 + 1))x) + 4 &= ((-2 + 2)x) + (-1); && \text{Definition of 3.} \\
(((-2 + 2) + 1)x) + 4 &= ((-2 + 2)x) + (-1); && \text{Associativity of } +. \\
((0 + 1) \cdot x) + 4 &= (0 \cdot x) + (-1); && \text{Definitin of } -2 \text{ (used twice).} \\
(1 \cdot x) + 4 &= (0 \cdot x) + (-1); && \text{Definitin of 0.} \\
x + 4 &= 0 + (-1); && \text{Identity of } \times \text{ and } 0 \cdot x = 0 \text{ proved above.} \\
x + 4 &= -1; && \text{Definition of 0.} \\
(x + 4) + (-4) &= -1 + (-4); && \text{Equality is a congruence for } +. \\
x + (4 + (-4)) &= -1 \cdot 1 + (-1)4; && \text{Associativity \& property above.} \\
x + 0 &= -1(1 + 4); && \text{Definition of } -4. \\
x &= -1 \cdot 5. && \text{Definition of 0 \& def. of 5.} \\
x &= -5 && \text{Property above.}
\end{aligned}$$

□

1.2 Solve for x in the following equation making sure to indicate at **each** step what axioms must be used. Do one step at a time and do not skip steps.

$$6x - 9 = x^2.$$

Despite being extremely tedious, showing steps does have merit. To illustrate this consider the following example.

1.3 Find the error in the following seemingly routine argument.

$$\begin{aligned} x &= 2; \\ x^2 &= 2x; \\ x^2 - 4 &= 2x - 4; \\ (x-2)(x+2) &= 2(x-2); \\ x + 2 &= 2; \\ 4 &= 2. \end{aligned}$$

1.5 Varieties

Most of the axioms we encounter in algebra involve equations between formulas in the operations we are intending to study. For example, in the integers the commutative law is presented as $x + y = y + x$, for all $x, y \in \mathbb{Z}$. *Universal Algebra* is the branch of algebra that tries to study everything algebraic by considering specific examples as solutions to particular equations. In that way we might hope to discover universal properties and apply them uniformly to all our systems of numbers without ever consider the details of an example.

Recall that the signature σ is a list of operator types that can be applied to a set S. For example, the integers $\mathbb{Z}$ have a binary operation '$+$' of addition, a binary operation '$\cdot$' for multiplication, a unary operation '$-$' to take negatives, a nullary operator 0 – the additive identity, and a nullary operator 1 – the multiplicative identity. So the signature that captures all these common properties would be $\sigma = \{+, \cdot, -, 0, 1\}$. The important aspect to remember is that signatures do not carry the information of the set S, they just represent the types of operations needed. Thus, the signature $\sigma = \{+, \cdot, -, 0, 1\}$ applies equally well to $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. This signature also applies to $\mathbb{N}$, but because there is no natural way describe negatives in 0 in $\mathbb{N}$, their we simply assume the associated operations are defined nonwhere (recall operations are partial function so their domain can be empty).

Definition 1.5.1. Fix a signature σ.

(i) A *σ-formula* is a conjunction of variable and operations in σ.

(ii) A *σ-equation* is sentence of the form $\phi(x_1, \ldots, x_\ell) = \gamma(x_1, \ldots, x_\ell)$ where ϕ and γ are σ-formulas in variables $x_1, \ldots, x_\ell$.

When σ is obvious from context we omit including it in the notation.

Example 1.5.2. For the signature $\sigma = \{+, \cdot, -, 0, 1\}$, the following are σ-formulas $x + y$, $x \cdot y + 1$, and $x \cdot (y - z)$. Also, $x + 2$ and x^2 are not formulas because 2 is not in the signature.

Definition 1.5.3. For a signature σ and a set Φ of σ-equations, the *σ-variety* $\mathfrak{V}(\Phi)$, with laws Φ is the class of all models S such that for all equations $\phi(x_1, \ldots, x_\ell) = \gamma(x_1, \ldots, x_\ell)$ in Φ it follows that

$$\forall s_1, \ldots, s_\ell \in S, \qquad \phi(s_1, \ldots, s_\ell) = \gamma(s_1, \ldots, s_\ell).$$

Example 1.5.4. (i) The variety $\mathfrak{V} = \mathfrak{V}(x \cdot 1 = x = 1 \cdot x)$ (where the implied signature is $\{\cdot, 1\}$) consists of all sets with a binary operation and an identity for that operation. Thus, $\langle \mathbb{N}, + \rangle \notin \mathfrak{V}$ because $\mathbb{N}$ does not have an identity under addition, but $\langle \mathbb{N}, \cdot \rangle$ and $\langle \mathbb{Z}, + \rangle$ are in $\mathfrak{V}$.

(ii) The variety $\mathfrak{V} = \mathfrak{V}(x \cdot x = x)$ (where the implied signature is $\{\cdot\}$) consists of all sets with a binary operation which is *idempotent*, that is, $x^2 = x$. For example, in the rational numbers, $x \star y = \frac{1}{2}(x + y)$ has the property that $x \star x = x$, so $\langle \mathbb{Q}, \star \rangle \in \mathfrak{V}$. However, $\langle \mathbb{Q}, \cdot \rangle$ is not in $\mathfrak{V}$ because $2 \cdot 2 = 4 \neq 2$.

Later we will give a detailed description of the many useful varieties we encounter in algebra. For now we recommend some exploration based solely on the definitions. Notice that for some products in the exercise below the associative identity is replaced by some non-obvious rules.

1.4 Decide if the following varieties contain the candidate indicated.

(a) $\langle \mathbb{Q}, \cdot \rangle \in \mathfrak{V}(x \cdot y = y \cdot x)$.

(b) $\langle M_2(\mathbb{Q}), \cdot \rangle \in \mathfrak{V}(x \cdot y = y \cdot x)$.

(c) $\langle M_2(\mathbb{Q}), \bullet \rangle \in \mathfrak{V}(x \cdot (y \cdot z) = (x \cdot y) \cdot z,\ x \cdot y = y \cdot x)$ where $X \bullet Y = \frac{1}{2}(XY + YX)$.

(d) $\langle M_2(\mathbb{Q}), \bullet \rangle \in \mathfrak{V}((x \cdot x) \cdot (y \cdot x) = ((x \cdot x) \cdot y) \cdot x)$ where $X \bullet Y = \frac{1}{2}(XY + YX)$.

(e) $\langle M_2(\mathbb{Q}), [,] \rangle \in \mathfrak{V}(x \cdot (y \cdot z) = (x \cdot y) \cdot z)$ where $[X, Y] = XY - YX$.

(f) $\langle M_2(\mathbb{Q}), [,] \rangle \in \mathfrak{V}(x \cdot x = 0,\ x \cdot (y \cdot z) = (x \cdot y) \cdot z + y \cdot (x \cdot z))$ where $[X, Y] = XY - YX$.

1.6 Standard Models

1.6.1 Semigroups, Monoids, & Groups

Further Reading Jacobson §1.1

The claims in Example-1.4.8 leaped a bit ahead and assumed things of the natural numbers which, however evident, are not the exact meaning of counting. These do need some clarification. For example, it is the definition of multiplication that $m \cdot n = \overbrace{n + \cdots + n}^{m}$, but that sum is not entirely well-defined because addition is a binary operation and so there is a need to include parentheses.

Further Reading Jacobson §0.7.

The justification is obvious, addition of natural number satisfies the is associative axiom (below). However, we will not prove this point either.

Variety of Semigroups

A binary operation $\star$ on a set S is *associative* when

$$s \star (t \star u) = (s \star t) \star u \qquad (\forall s, t, u \in S).$$

We call $\langle S, \star \rangle$ (or sometimes just S) a *semigroup*.

As the name semigroups suggests, number systems that model only the associative axiom are only part of what we hope for in a useful number system. Many semigroups also model the identity axiom which makes for a powerful combination. Thus, models for both the identity and associative axioms are given a name of their own; they are *monoids*.

Example 1.6.1. The natural numbers are a semigroup under addition and a monoid under multiplication.

A much more general example of semigroups and even monoids is provided by functions. The reason is the following well-known observation.

Lemma 1.6.2. *Function composition, whenever defined, is associative.*

Proof. Let $f : A \to B$, $g : B \to C$, and $h : C \to D$ be functions. For each $x \in A$,

$$\begin{aligned}((h \circ g) \circ f)(x) &= (h \circ g)(f(x)) = h(g(f(x)) \\ &= h((g \circ f)(x)) = (h \circ (g \circ f))(x).\end{aligned}$$

Therefore $(h \circ g) \circ f = h \circ (g \circ f)$. □

Definition 1.6.3. Fix a set X. The *transformation monoid* on X is the set of all functions on X with composition as the binary operation and the identity function as the identity of the monoid.

Even though semigroups are not the most robust numbers we might hope to discover, they are amongst the most common because they have such a low threshold to be satisfied – one axiom. More surprising, they are equivalent to the study of the most primitive computational circuits, so called finite state automata. Those circuits do not require memory and they control the bulk of all machines such as elevators and microwave ovens. It may appear that these systems require memory to lock in the floor your want to reach or the time to heat the water. But in fact the range of possible outcomes is so narrow that the machine simply takes the instruction and begins running, it does not need to circle back and recall what input you gave it to complete its task. Because of this tremendous utility, modern studies of semigroups in mathematics have increased considerably.

The natural numbers are not as robust as we hope. Of course we cannot use them to solve $x + 5 = 2$ nor $5x = 2$. For that we need negative numbers and fractions. Before making models of integers and rational numbers let us absorb the property of being a 'negative'.

Variety of Groups

For a monoid $\langle M, \cdot \rangle$, an *inverse* to an element $m \in M$ is an element m' such that

$$m \cdot m^{-1} = 1 = m^{-1} \cdot m.$$

A *group* is a monoid where every element has an inverse.

Again we emphasize that axioms are written with variables for both sets and numbers as well as operations. So in the inverse axiom we wrote m^{-1} because it is memorable. However, when we work with the integers under addition it is clear that the inverse of 2 is -2 and not 2^{-1} – which is not an integer. So the unary operator of inverse in the axiom must be replaced with the unary operation for the specific model, in the last case, negation.

Example 1.6.4. (a) The natural numbers are not a group under addition nor multiplication. However, the set $\{1\}$ is a group under multiplication.

(b) The integers are a group under addition but not under multiplication.

(c) The rational numbers are a group under addition but not under multiplication (because we do not have an inverse for 0).

1.5 Alternate Group Axioms In this exercise we show that the same theory can be specified by different axioms. The example is the theory of groups with the axioms as described above. Now show that every model $\langle G, \cdot \rangle$ satisfying the group axioms also satisfies the following sentence.

(a) $\cdot$ is associative

(b) There is an $e \in G$ such that for all $g \in G$, $e \cdot g = g$.

(c) For every $g \in G$ there is a $g' \in G$ such that $g \cdot g' = e$.

Next prove that every model $\langle G, \cdot \rangle$ satisfying the axioms (a), (b), & (c) also satisfy the group axioms. Thus, the two theories are the same.

1.6 Inverse products Suppose $\langle G, \cdot \rangle$ is a group. Define $\langle G, \div \rangle$ as the set with binary operation

$$g \div h = g \cdot h^{-1} (\forall g, h \in G).$$

Prove that $\langle G, \div \rangle$ is a group then every element $g \in G$ satisfies $g^2 = 1$. Otherwise show that $\langle G, \div \rangle$ does not satisfy *any* of the usual axioms of a group. Notice this explains how subtraction and division behave so differently from addition and multiplication.

Permutation groups

Definition 1.6.5. For a set X, the *symmetric group* on X, denoted Sym(X) is the set of all invertible functions $X \to X$. Also, invertible functions on a set are referred to as *permutations* of X. If $X = \{1, 2, \ldots, n\}$ then often Sym(X) is abbreviated S_n. A group consisting of permutations and whose operation is composition is called a *permutation group*.

We will later encounter Cayley's Theorem which explains that every monoid can be expressed as a transformation monoid (with some loss of information) and every group can be expressed as a group of permutation (without any loss of information). For now we pause to explain some of the common methods to describe permutations using a method known as *cycle notation*.

A function $f : \{1, 2, 3, 4\} \to \{a, b, c, d\}$ must assign a single value to each input 1, 2, 3, and 4. This can be captured by a table whose first row is the inputs and whose second row is the assigned output. For example:

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ b & a & c & a \end{pmatrix}.$$

We interpret this to mean $f(1) = b$, $f(2) = a$, $f(3) = c$, and $f(4) = a$ so this determines a function. When f is a transformation, i.e. a function from a set X back into X, then two natural methods to encode f present themselves.

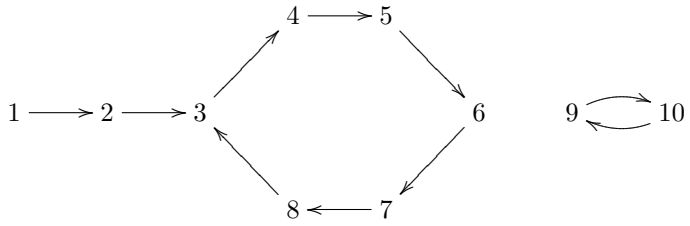
The most common is what we know as the *graph* of f. There we plot f as the points $(x, f(x))$ for every x in the domain of f. For instance, $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2$ instructs us to plot the parabola $(x, f(x))$. Without question this technique is very powerful, e.g. with it we discover the notion of tangents to curves and so construct the Calculus.

When X is a small set, especially when X is finite, the power of graphs is less obvious because the graphs of functions $f : X \to X$ do not take on familiar shapes. (In actuality, functions $f : \mathbb{R} \to \mathbb{R}$ are also pointless to draw in general because a random function of that kind would not be continuous and so it would appear as random dots spread around the plane.) However, because f starts

and ends in X we might look for eventual patterns. For instance, we start with $x_0 \in \mathrm{X}$ and find $f(x_0) = x_1 \in \mathrm{X}$. Next we try $f(f(x_0)) = f(x_1) = x_2 \in \mathrm{X}$, etc. This produces an infinite sequence of element in X. When the size, which we denote by $|\mathrm{X}|$, of X is finite it follows by the *pigeon-hole-principle*[5] that at some point

$$x_{i+1} \in \{x_0, x_1 = f(x_0), x_2 = f^2(x_0), \ldots, f^i(x_0)\}.$$

At that point there is no reason go further because once $f^{i+1}(x_0) = f^j(x_0)$ then also $f^{i+2}(x_0) = f^{j+1}(x_0)$, etc. and we entire an infinite repeating cycle. Graphically this might appear as follows.



This would represent the function:

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 3 & 10 & 9 \end{pmatrix}.$$

Notice f is not invertible as there is no output equal to 1. This gives us a visual test of invertibility. If we draw the diagram as above, the function is invertible if, and only if, it is a collection of cycles. Notice we can record cyclces in a simply manner. Instead of drawing a table inputs to outputs, we can simply record the sequence $(x_0, f(x_0), f^2(x_0), \ldots, f^i(x), \ldots)$. If f is a permutation then it is enough to record all the cycles, and we recognize no cycle in f has a value from any other cycle. So we say that f is the 'product of disjoint cycles'. This needs a little formality.

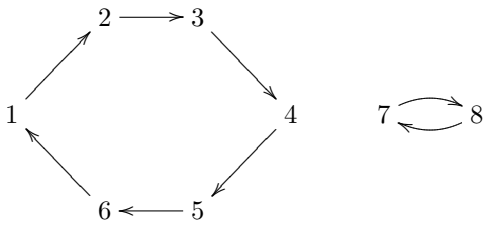
Definition 1.6.6. Let X be a finite set and $f : \mathrm{X} \to \mathrm{X}$ a permutation of X.

(a) A *base* for f is subset $\mathrm{Y} \subseteq \mathrm{X}$ such that for every $y \in \mathrm{Y}$ and every $n \in \mathbb{Z}$, $f^i(y) \in \mathrm{Y}$ implies that $y = f^i(y)$.

(b) A cycle of f is $(y, f(y), f^2(y), \ldots, f^i(y), \ldots)$.

.

For example, the following two disjoint cycles identify a permutation of $\{1, \ldots, 8\}$



A base for this permutation would be $\{1, 7\}$, also $\{2, 8\}$ and $\{1, 8\}$ are bases. Yet $\{1, 2, 7\}$ and $\{1, 3\}$ are not bases. The cycles of this permutation are $(1, 2, 3, 4, 5, 6)$ and $(7, 8)$, so we recored the entire permutation as $(1, 2, 3, 4, 5, 6)(7, 8)$.

ing Jacobson

1.6.2 Rings & fields

The natural numbers have two binary operations which are strongly related so there should be an axiom considered that explains at least a part of this relationship. This is the distributive axiom, recall its definition. For a set S with binary operations $\cdot$ and $+$ (again these are the usual signatures but not necessarily the actual labels of the operations),

$$\begin{aligned} s(t+u) &= (st)+(su), \\ (s+t)u &= (su)+(tu) \end{aligned} \qquad (\forall s,t,u \in S).$$

We say that $\cdot$ *distributes* over $+$.

1.7 Describe all the models satisfying the following axioms: for a set S

(i) $\langle S, + \rangle$ is a group,

(ii) $\langle S, \cdot \rangle$ is a group, and

(iii) $\cdot$ distributes over $+$.

[Hint: notice a model for these axioms would represent division by zero.]

Theorem 1.6.7. *Suppose $\langle S, + \rangle$ has an identity 0 and $\langle S, \cdot \rangle$ an identity 1. If $\cdot$ distributes over $+$ then for all $s \in S$, $s \cdot 0 = 0 = 0 \cdot s$.*

Proof. Let 0 be the identity of $+$ and 1 the identity of $\cdot$.

$$\begin{aligned} 1 &= 1+0 \\ 1 \cdot s &= (1+0) \cdot s \qquad (\forall s \in S) \\ s &= (1 \cdot s) + (0 \cdot s) \\ s &= s + (0 \cdot s). \end{aligned}$$

Thus, for all $s \in S$, $s + (0 \cdot s) = s$ and by beginning with $1 = 0 + 1$ and repeating the steps above we find $(0 \cdot s) + s = s$. Therefore $(0 + s)$ satisfies the identity axiom for $\langle S, + \rangle$. But we saw in Theorem-1.4.1 that identities are unique. Therefore $0 \cdot s = 0$. The argument can be adapted to show $s \cdot 0 = 0$ as well. □

Definition 1.6.8. A *ring* is a $\{\cdot, +, -, 0\}$-algebra R where

(i) $\langle R, +, -, 0 \rangle$ is an abelian group,

(ii) $\langle R, \cdot \rangle$ is a semigroup, and

(iii) $\cdot$ distributes over $+$.

If R has a multiplicative identity 1 then say R is a *unital ring*. If the multiplication in R is commutative we say R is a *commutative ring*.

Example 1.6.9. (i) $2\mathbb{Z}$ is a ring but not unital.

(ii) $M_2(\mathbb{Q})$ is a unital ring but not a commutative unital ring.

(iii) $\mathbb{Z}$ is a commutative unital ring.

(iv) The set $\{0\}$ with the trivial operations $0 + 0 = 0$ and $0 \cdot 0 = 0$ is a commutative unital ring in which $1 = 0$. Some authors decline to allow such things to be rings. In any case such a thing is called a *trivial ring*.

[5] If you have n pigeon-holes and $n+1$ pigeons, then two of them must share a home.

(v) $M_2(\mathbb{Q})$ with product $x \circ y = \frac{1}{2}(xy+yx)$ is not a ring because the multiplication is associative. (However, the multiplication is unitl and commutative.)

Remark 1.6.10. It is not uncommon to find authors insist that all rings be unital. This is acceptable but has a few delicate problems. First of all, when authors write this they often implicitly (or sometimes explicitly) require that $0 \neq 1$. We have seen that if $0 = 1$ then every element in the ring equals 0, i.e. the ring is $\{0\}$. However, insisting that $0 \neq 1$ is *not* and equation. So this is not something we can impose as a law of varieties. Indeed we will see later that varieties must be closed to subalgebras and $\{0\}$ is always a subalgebra of a ring. Therefore we cannot make a variety out of the assumption that all rings are unital where $0 \neq 1$. Instead, if we insist that all rings are unital we must permit the silly instance of the ring $\{0\}$.

1.6.3 Modular rings

Think of a clock. Times of day are recorded as numbers $1, 2, \ldots, 12$ with the agreed rule that $x + y$ will be reduced back to a number between 1 and 12, e.g. 5-o-clock plus 9-o-clock is 2-o-clock rather than 14-o-clock. The precise formula involves the division algorithm.

Theorem 1.6.11 (Division Algorithm)**.** *For every pair (n, m) of integers where $m \neq 0$, there exists unique integers q (the quotient) and r (the remainder) such that $0 \leq r < q$ and $n = mq + r$.*

Proof. Suppose that $n \geq 0$. For every positive integer q, as $m \neq 0$, $mq < mq + 1 < m(q + 1)$. Therefore the sequence $0 < m < 2m < \cdots$ eventually is larger than n. Let q be the smallest nonnegative integer such that $mq \leq n < m(q+1)$.[6] Hence, $0 \leq r = n - mq < m$ and $n = mq + r$. If $n < 0$ then there are unique integers q and r such that $-n = mq + r$ and $0 \leq r < m$. So

$$n = -mq - r = -mq - m + m - r = m(-q - 1) + (m - r).$$

Notice $(-q - 1)$ and $(m - r)$ are integers and $0 \leq m - r < m$. □

Definition 1.6.12. Fix an integer m. Let $\mathbb{Z}/m = \{0, 1, 2, \ldots, m-1\}$. For each $x, y \in \mathbb{Z}/n\mathbb{Z}$ define

$$\begin{aligned} x + y &\equiv r \pmod{m} && \text{if } x + y = mq + r, 0 \leq r < m; \\ xy &\equiv r' \pmod{m} && \text{if } xy = mq' + r', 0 \leq r' < m. \end{aligned}$$

We call this addition and multiplication *modulo* m.

It is safest to draw a distinction between regular addition and addition modulo q. This is why we write the symbol $\equiv$ instead of $=$, e.g. to avoid confusing $5 + 9 \equiv 2 \pmod{q}$ with writing $5 + 9 = 14$.

Theorem 1.6.13. *For every $n \in \mathbb{Z}$, $\mathbb{Z}/n$ is a commutative unital ring. Notice $\mathbb{Z}/n$ is isomorphic to $\mathbb{Z}/(-n)$, $\mathbb{Z}/1$ a ring with just one element $\{0\}$ and $\mathbb{Z}/0$ is the ring $\mathbb{Z}$.*

1.6.4 Division rings & Fields

Definition 1.6.14. (i) A *division ring* is a unital ring D along with the following property:

$$\forall x \in D, x \neq 0 \Rightarrow \exists x^{-1} \in D, xx^{-1} = 1. \tag{1.1}$$

[6] We are implicitly using the "well-ordering" principle of nonnegative integers. This is equivalent to induction. To see the correspondence consider Halmos' *Naïve Set Theory*.

(ii) A *field* is a commutative division ring.

Fields are quite common by by this point. Examples include $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. Less well-known examples include $\mathbb{Z}/p$ for a prime p.

Lemma 1.6.15. *In a non-trivial unital ring R, if $x, y \in R$ such that $xy = 0$ then either x or y has no inverse.*

Proof. Let $x, y \in R$ such that $xy = 0$. Suppose that x has an inverse x^{-1}. It follows that $0 = x^{-1}0 = x^{-1}(xy) = (x^{-1}x)y = 1y = y$. Therefore $y = 0$ and so y has no inverse. On the other hand if x has no inverse then we also have proved the lemma. So in all cases either x or y has no inverse. □

Notice this lemmas does *not* say that either x or y is 0!. Indeed, consider matrices:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Theorem 1.6.16. *For every $n > 1$, $\mathbb{Z}/n$ is a field if, and only if, $n = \pm p$ where p is a prime.*

Proof. Let $n = ab$ where $a, b \in \mathbb{Z}$ and $a, b > 0$. Then $ab \equiv n \equiv 0 \pmod{n}$. By Lemma-1.6.15, this implies that either a or b has no inverse.

In particular, if $\mathbb{Z}/n$ is a field then the only number without an inverse is 0 and so either $a \equiv 0 \pmod{n}$ or $b \equiv 0 \pmod{n}$. If $a \equiv 0 \pmod{n}$ then n divides $a - 0 = a$. Yet $a, n > 0$ and $n = ab$ so $n = a$ and $b = 1$. As this is done for arbitrary factorization of n it follows that if $\mathbb{Z}/n\mathbb{Z}$ is a field then n is prime.

On the other-hand, if p is not a prime then there exists $n > a, b > 1$ such that $n = ab$. Thus, neither a nor b is divisible by n and so neither a nor b is equivalent to 0 mod n. As $ab \equiv 0 \pmod{n}$ we have two nonzero elements, $a + n\mathbb{Z}$ and $b + n\mathbb{Z}$ in $\mathbb{Z}/n$ which have no inverses. We conclude $\mathbb{Z}/n$ is not a field. □

Remark 1.6.17. It is correct to regard $\mathbb{Z}/1\mathbb{Z}$ as a field of size 1. However this field is isomorphic to the trivial ring $\{0\}$ and that does require some delicacy when working. For example, a vector space over $\{0\}$ is not exactly obvious to describe. The use of the field of size 1 seems to have begun with the French mathematician Jacques Tits (who has received both a Fields medal and the Abel prize). It has also been adopted by Field medal winner Alain Connes and several others.[7] Using this field properly can be difficult and modern texts on algebra still refrain from including this interesting case.

Remark 1.6.18. The definition of a field (or more generally a division ring) may suggest that the class of fields makes a variety because it is defined by equations. However, the inverse axiom for multiplication applies only to non-zero elements of F. Thus it is *not* true that F satisfies an equation $xx^{-1} = 1$ because it is only true of *some* of the elements of F. However, this does not prove the class of fields from being a variety by means of some other perhaps more clever system of equations. However, we will give a very clear reason why this cannot be the case later in the section on direct products.

[7] Watch Alain Connes speak about this field on Youtube: "Fun with F_1".

Algebraic Extensions

In Chapter 1 we proved that $\sqrt{2}$ is not a rational number. Thus, we were motivated to invent new numbers in such a way as to include a model which would posses a value α where $\alpha^2 = 2$. We mentioned briefly that this can be done without constructing all decimal numbers. In this section we expand on that construction.

Suppose we want a new number α to be extend $\mathbb{Q}$ in such a way that we can add and multiply with α and all the usual fractions. This requires that we include products and sums such as:

$$3\alpha, \quad \alpha^2, \quad 2+\alpha, \quad -\frac{1}{2}+\frac{\alpha}{7}+\alpha^{2011}.$$

Now if our goal is to make α behave like a squareroot of 2 then we would insist that $\alpha^2 = 2$. We should resist the temptation at this point to write $\alpha = \sqrt{2} = 1.41\cdots$. Indeed, we are not even concerned with which of the two possible roots of $x^2 - 2$ we might pick to become α. Notice already the assumption $\alpha^2 = 2$ implies we no longer need α^2, α^3, α^4, α^5, etc. because we could always rewrite these as 2, 2α, 4, 4α, etc. So, a number system that extends $\mathbb{Q}$ to include a squareroot of 2 would only need to add in an α along with all the following numbers:

$$\frac{a}{b}+\frac{c}{d}\alpha \qquad\qquad (\forall a/b, c/d \in \mathbb{Q}).$$

So we define

$$\mathbb{Q}[\alpha] = \{x + y\alpha : x, y \in \mathbb{Q}\}.$$

Unfortunately this notation does not quite make sense yet. To start out, what is meant by $x + y\alpha$ if we do not yet have a new number α? The answer is that α is just a place holder for a location as follows:

$$x + y\alpha = (x, y) \in \mathbb{Q} \times \mathbb{Q}. \tag{1.2}$$

So as sets $\mathbb{Q}[\alpha] = \mathbb{Q} \times \mathbb{Q}$. The notation we pick will become clear once we consider operations on this set. First we add.

$$(x + y\alpha) + (z + w\alpha) = (x + z) + (y + w)\alpha \qquad (\forall x + y\alpha, z + w\alpha \in \mathbb{Q}[\alpha])$$

Notice in coordinates this reads:

$$(x, y) + (z, w) = (x + z, y + w)$$

so the addition we have is the same addtion we expect for vectors. Indeed, we treat $\mathbb{Q}[\alpha]$ as 2-dimensional vector space over $\mathbb{Q}$ with a basis of $\{1, \alpha\}$ (technically $\{(1, 0), (0, 1)\}$). Notice that as vectors

$$(x, y) = (x, 0) + (0, y) = x(1, 0) + y(0, 1) = x \cdot 1 + y \cdot \alpha = x + y\alpha.$$

That justifies the notation we selected. This convention should look familiar if you have encountered complex numbers before. Specifically complex numbers are written as $a + bi$ where $a, b \in \mathbb{R}$ but the are also treated as the coordinate (a, b) in the xy-plane. The only difference here is that we are using $\mathbb{Q}$ instead of $\mathbb{R}$ and $\alpha^2 = 2$ rather than $i^2 = -1$. One crucial fact to remember is that $\alpha = (0, 1)$ and so $\alpha^2 = (2, 0)$ – this is not the same as $(0, 2) = 2\alpha$.

The multiplication on $\mathbb{Q}[\alpha]$ is created by mimicing the properties of the distributive property which we hope will hold in order to create a ring structure on $\mathbb{Q}[\alpha]$.

$$\begin{aligned}(a+b\alpha)(c+d\alpha) &= a(c+d\alpha)+bi(c+d\alpha) \\ &= ac+a(d\alpha)+(b\alpha)c+(b\alpha)(d\alpha) \\ &= ac+(ad)\alpha+(bc)\alpha+(bd)\alpha^2 \\ &= (ac+2bd)+(ad+bc)\alpha.\end{aligned} \tag{1.3}$$

Notice that our method to multiply these number is forced on us by the distributive property along with the goal of making $\alpha^2 = 2$. We also used the associativity of multiplication in $\mathbb{Q}$ along the way and allowed ourselves to commute rational numbers passed α. We mention this because in subsequent constructions these assumptions are crucial.

To make the distributive property very noticeable it can help to draw the multiplication in a table where the product is illustrated as 'length times width' as follows:

$\cdot$	c	$d\alpha$
a	ac	$ad\alpha$
$b\alpha$	$bc\alpha$	$2bd$

Once we combine the terms in the grid we recover the product we had above.

Quaternions

Further reading Jacobson §2.4

We close with an example a division ring which is not a field. This example is quite popular and indeed inspired its creator to vandalize a bridge by inscribe the rules of the multiplication into the bridge. Subsequently they are often referred to as the *Hamiltonians* though it is equally common to see the listed as *Quaternions*.

Let $\mathbb{H} = \mathbb{C} \times \mathbb{C}$ as a set. However, owing to our eventual use we choose to describe the elements of $\mathbb{H}$ not as ordered pairs (x, y), where $x, y \in \mathbb{C}$, but instead we write $x + y\hat{\jmath}$ (similar to complex numbers which we write as $a + bi$ where $a, b \in \mathbb{R}$). Use the following operations:

$$(x + y\hat{j}) + (z + w\hat{\jmath}) = (x + z) + (y + w)\hat{\jmath} \quad (\forall (a + b\hat{\jmath}, c + d\hat{\jmath} \in \mathbb{H} = \mathbb{C} \times \mathbb{C}).$$

The use of $\hat{\jmath}$ is simply because we continue on from i. Notice we have numbers such as $i+2\hat{\jmath}$ and $i\hat{j}$ in $\mathbb{H}$. To help understand where in the term each number is we write $\hat{\imath}$ instead of $i + 0\hat{\jmath}$. Thus, $\hat{\imath} + \hat{\jmath}$ is the number $(1, 1) \in \mathbb{C} \times \mathbb{C} = \mathbb{H}$ and $i\hat{j}$ means the number $(0, i) \in \mathbb{C} \times \mathbb{C} = \mathbb{H}$. Once we describe the multiplication in $\mathbb{H}$ we will discover what we hoped would be true, that $i\hat{j} = \hat{\imath}\hat{j}$ – in coordinates that would read as $(1, 0) \cdot (0, 1) = (0, i)$.

Next we consider the multiplication. We follow the example of complex numbers. Notice that our method to multiply complex number is forced on us by the distributive property and the goal of making $i = \sqrt{-1}$. For each $a + bi, c + di \in \mathbb{C} = \{x + yi : x, y \in \mathbb{R}\}$ we define:

$$(a+bi)(c+di) = a(c+di)+bi(c+di) = ac+a(di)+(bi)c+(bi)(di) \tag{1.4}$$

$$= (ac-bd)+(ad+bc)i. \tag{1.5}$$

Notice we also used the associativity of multiplication in $\mathbb{R}$ and allowed ourselves to commute numbers in $\mathbb{R}$ passed i. We mention this because in subsequent constructions these assumptions are crucial. To make the distributive property

very noticeable it can help to draw the multiplication in a table where the product is illustrated as 'length times width' as follows:

$\cdot$	c	di
a	ac	adi
bi	bci	$-bd$

Once we combine the terms in the grid we recover the product we had in (1.4).

Now we return to $\mathbb{H}$ and describe a product for this set which behaves almost the same as in $\mathbb{C}$. Recall that for a complex number $z = a + b\hat{\imath}$, $\bar{z} = a - b\hat{\imath}$. With that in mind we can describe the multiplication table for $\mathbb{H}$.

$\cdot$	z	$w\hat{\jmath}$
x	xz	xw
$y\hat{\jmath}$	$\bar{y}z$	$-\bar{y}w$

Notice in this product we encounter $\hat{\imath}^2$, which is still a complex number and so as before $\hat{\imath}^2 = -1$ (in the multiplication table this corresponds to $x = \hat{\imath}$, $y = 0$, $z = \hat{\imath}$, and $w = 0$. We also have $\hat{\jmath}^2$ which we declare to be -1 so that $\{a + b\hat{\jmath} : a, b \in \mathbb{R}\}$ is another copy of $\mathbb{C}$. We see that reflected in the multiplication table, i.e. using $x = 0$, $y = 1$, $z = 0$, and $w = 0$. And finally we encounter $\hat{\imath}\hat{\jmath}$ which is in our multiplication table ($x = \hat{\imath}$, $y = 0$, $z = 0$, $w = 1$). This is assigned the value $\hat{\imath}\hat{\jmath}$ – nothing else.

Theorem 1.6.19. *$\mathbb{H}$ is a division ring that is not a field.*

1.6.5 Jordan & Lie rings

Further reading Jacobson §7.5

Definition 1.6.20. A *nonassociative ring* is a $\{\bullet, +, -, 0\}$-algebra A (where $\bullet$ is a binary operation) such that

(i) $\langle A, +, -, 0\rangle$ is an abelian group,

(ii) $\bullet$ distributes over $\cdot$.

It would also be better to have written "nonassociative-ring" so that one is not lead to missundertand that the algebra A is first and formost a ring – *it is not usually a ring!* Evidently every ring is also a nonassociative ring, but the converse is false. So it is perhaps more descriptive to regard "nonassociative-ring" to mean "not necessarily associative ring". However inadequate, the name has stuck.

Definition 1.6.21. A *Lie ring* is a $\{[,], +, -, 0\}$-algebra L (where $[,]$ is a binary operation) which is a nonassociative ring satisfying

(i) $[x, x] = 0$,

(ii) (Jacobi Identity) $[x, [y, z]] = [[x, y], z] + [y, [x, z]]$.

As before, a Lie ring is not usually a ring and so it would have been best to title these as "Lie-rings". Yet, as with many things through the passage of time this unfortunate confusion has not been rectified.

Example 1.6.22. (i) For every commutative ring K (e.g. $K = \mathbb{R}$), the set $M_n(K)$ equiped with it usual addition and with the product $[X, Y] = XY - YX$ makes a Lie ring. It may seem silly to study the ring $M_n(K)$ using a clunky nonassociative product such as $[X, Y]$. However, the next example demonstrates that sometimes this product is the only obvious product.

(ii) For every commutative ring K (e.g. $K = \mathbb{R}$), the set

$$\mathrm{Alt}_n(K) = \{X \in M_n(K) : X^t = -X\}$$

equiped with it usual addition and with the product $[X, Y] = XY - YX$ makes a Lie ring. Notice $\mathrm{Alt}_n(K)$ does not make a ring in the usual way because, for some $X, Y \in \mathrm{Alt}_n(K)$, we can have

$$(XY)^t = Y^t X^t = (-Y)(-X) = YX \neq -(XY).$$

In particular $XY \notin \mathrm{Alt}_n(K)$.

Lemma 1.6.23. *A Lie ring $L \neq \{0\}$ then L cannot be unital. Furthermore, for every $x, y \in L$, $[x, y] = -[y, x]$ so we say that the multiplication is "skew-commutative."*

Proof. If L is unital then there is a number 1 such that for all $x \in L$, $[x, 1] = x$. However, $[1, 1] = 0$ by the first axiom and so this forces $1 = 0$ and so $L = \{0\}$.

Next, for all $x, y \in L$, by the first axiom, $[x + y, x + y] = 0$. We also know $[,]$ distributes over addition and so we have the following.

$$\begin{aligned} 0 &= [x + y, x + y] = [x, x + y] + [y, x + y] \\ &= [x, x] + [x, y] + [y, x] + [y, y] \\ &= [x, y] + [y, x]. \end{aligned}$$

□

Remark 1.6.24. The Jacobi identity may not appear easy to remember but in fact it is something quite familiar. Recall the Leibniz rule reads:

$$\frac{d}{dx}(f \cdot g) = \left(\frac{d}{dx} f\right) g + f \left(\frac{d}{dx} g\right). \tag{1.6}$$

If we think of $\frac{d}{dx}$ as x, f as y, and g as z then this formula is nothing more than example of the Jacobi identity.

Definition 1.6.25. A *Jordan ring* is a $\{\circ, +, -, 0\}$-algebra J (where $\circ$ is a binary operation) which is a nonassociative ring satisfyingA *field* is a commutative unital ring F along with the following property:

$$\forall x \in F, x \neq 0 \Rightarrow \exists x^{-1} \in F, x x^{-1} = 1. \tag{1.7}$$

(i) $x \circ y = y \circ x$,

(ii) $(x \circ x) \circ (y \circ x) = ((x \circ x) \circ y) \circ x$.

If J has multiplicative identity then we say J is *unital*.

Similar to Lie rings, Jordan rings are not usually rings.

Example 1.6.26. (i) For every commutative ring K (e.g. $K = \mathbb{R}$), the set $M_n(K)$ equiped with it usual addition and with the product $X \circ Y = XY + YX$ makes a Jordan ring. Actually it is most common to assume K is has a $1/2$ and so we set $X \circ Y = \frac{1}{2}(XY + YX)$.

(ii) For every commutative ring K (e.g. $K = \mathbb{R}$), the set

$$\mathrm{Sym}_n(K) = \{X \in M_n(K) : X^t = X\}$$

equiped with it usual addition and with the product $X \circ Y = XY + YX$ makes a Jordan ring. Notice $\mathrm{Sym}_n(K)$ does not make a ring in the usual way because, for some $X, Y \in \mathrm{Sym}_n(K)$, we can have

$$(XY)^t = Y^t X^t = (-Y)(-X) = YX \neq XY.$$

In particular $XY \notin \mathrm{Sym}_n(K)$.

Further reading Jacobson §7.6

Altenative rings & Octonions

Definition 1.6.27. An *Alternative ring* is a $\{\cdot, +, -, 0\}$-algebra A which is a nonassociative ring satisfying

(i) (Alternative Laws) $x(xy) = (xx)y$ and $x(yy) = (xy)y$.

An alternative algebra with a 1 is called *unital*.

The name "alternative" relates to who the associative property partially holds. Define the *associator* as the formula:

$$[x, y, z] = (xy)z - x(yz). \tag{1.8}$$

Lemma 1.6.28. *(i) In a ring R, $[x, y, z] = 0$. In particular every ring is alternative.*

(ii) In an alternative ring R, $[x, y, z] = -[y, x, z] = [y, z, x] = [z, x, y] = -[z, y, x] = -[x, z, y]$. That is, if we rotate the letters we get the same associator but if we flip two letter we "alternate" signs.

Proof. (i). In a ring the multiplication is associative and so $[x, y, z] = (xy)z - x(yz) = 0$.

(ii). An an alternating ring $[x, x, y] = 0$ and $[x, y, y] = 0$. So we "polarize" these identities, which means to replace the letters with sums and use the distributive property to derive new identities.

$$\begin{aligned}
0 &= [x + y, x + y, z] = ((x + y)(x + y))z - (x + y)((x + y)z) \\
&= (xx + yx + xy + yy)z - (x + y)(xz + yz) \\
&= (xx)z + (yx)z + (xy)z + (yy)z - x(xz) - x(yz) - y(xz) - y(yz) \\
&= (xx)z - x(xz) + (yy)z - y(yz) + (xy)z - x(yz) + (yx)z - y(xz) \\
&= [x, y, z] + [y, x, z]
\end{aligned}$$

The other identite are seen by similar substitutions. □

We have shown that rings are alternative but it would not make sense to introduce alternative rings if there were not an important family of examples of alternative rings that are nonassociative, and hence not rings. There is such a family known as *Octonions*.

We repeat the process we used to create the quaternions but we make one more modification.

Let $\mathbb{O} = \mathbb{H} \times \mathbb{H}\hat{\ell}$ with addition of vectors, i.e.:

$$(a + b\hat{\ell}) + (c + d\hat{\ell}) = (a + c) + (b + d)\hat{\ell} \qquad (\forall a, b, c, d \in \mathbb{H}). \tag{1.9}$$

We multiply as follows

$\cdot$	z	$w\hat{\ell}$
x	xz	wx
$y\hat{\ell}$	$\bar{y}z$	$-w\bar{y}$

Theorem 1.6.29. *$\mathbb{O}$ is an alternative ring which is not associative.*

Definition 1.6.30. A nonassociative ring is a *semifield* if every element except 0 has an inverse.

Theorem 1.6.31. *$\mathbb{O}$ is a semifield that is not a field.*

1.7 Incompleteness and undecidability

We have now created a framework to study the method of reasoning as mathematics itself. The framework is largely algebraic as it consists of combining and substituting variables along with operations from logic (such as the binary operations of `and`, `or`, and the unary operations $\forall$, $\exists$, and `not`). So it makes sense to spend a small amount of time discovering some of the immediate algebraic implications of this approach. The result is fascinating and unexpected. For example, the word "consequence" suggests a reason (a proof) underpins the result. As a crude description, a proof of a sentence ϕ is a list $[\phi_1, \ldots, \phi_n]$ of sentences (in the same signature) in which $\phi = \phi_n$ and for each $1 \leq i < n$:

1. $\phi_i \in \Phi$ or
2. ϕ_i is the result of combining $[\phi_1, \ldots, \phi_{i-1}]$ by the rules of logic (substitution, inferences, the law of the excluded middle, etc.).

In 1930, as a response to a question of David Hilbert, Kurt Gödel gave the first demonstration of a sever restriction on proofs and truths.

Theorem 1.7.1 (Gödel's Incompleteness Theorem)**.** *Either, the theory of the integers has consequences that cannot be reached by a proof, or the theory of the integers has proofs reaching statements that are not consequences.*

The escence of Gödel's proof is as follows. We have seen that writing sentences can be done using symbols in a signature, symbols from logic, and variables. If we so wish we can give each of these finite symbols a number, say

- 1 stands for $+$,
- 2 represents $\cdot$,
- 3 replaces 1,
- 4 is used for $\forall$,
- 5 for (, and
- 6 for),
- 7 for $=$,
- etc.

If you are familiar with computer programs you might recognize that ASCII is an assignment of numbers to 256 common characters which is quite similar to Gödel's idea. This list ultimately stops once we have exhausted all the arithmetic and logic symbols we need. The actual number of symbols we need depends on how we think to write our logic. Say it stops at 99 (which is far larger than necessary). We still have not recorded variables. So we do this by assigning 100 to x_1, 101 to x_2, etc. so that we have an infinite number of variables should we need that. This means that every sentence we wish to write concerning the integers corresponds to a sequence of integers, for example,

$$\forall x_1(x_1 \cdot 1 = x_1)$$

might have appeared in our numbering as

$$(4, 100, 5, 100, 2, 3, 7, 100, 6, 0, \ldots).$$

We fill in the list with zeros to make it uniform, but we comment that each sequence has only finitely many nonzero values.

The next step is to convert lists of integers into one integer is such a way that we do not loose any information. That requires a one-to-one function. For that we rely on the Fundamental Theorem of Numbers.

The Fundamental Theorem of Numbers

Every natural number is a product of a unique list of primes.

In some sense it is the use of that theorem that highlights the need for a model as robust as the integers. That theorem is a indeed part of the Theory of Integers and so we may use it in creating our one-to-one function. We list the primes in order $\{p_1 = 2, p_2 = 3, p_4 = 5, \dots\}$. Then given a sequence $(a_1, \dots, a_n, \dots)$ of integers where only finitely many are non-zero, we create one integer as:

$$f(a_1, \dots, a_n, \dots) = p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n} \cdots . \tag{1.10}$$

This is a one-to-one function because of the Fundamental Theorem of Numbers. Thus, instead of talking about sentences P in logic we can talk about lists $A_P = (a_1, a_2, \dots)$ of integers with only finitely many nonzero entries. Next instead of talking A_p we now talk about a single positive integer $N_P = f(A_P)$. Hence, we can introduce a function which serves as a "lie-detector". That is, $g : \mathbb{N} \to \{0, 1\}$ has $g(n) = 1$ if there is a consequence P in the Theory of Integers such that $n = N_p$; otherwise $g(n) = 0$. For example,

$$\begin{aligned} P &\equiv \forall x_1 (x_1 \cdot 1 = x_1) \\ A_P &= (4, 100, 5, 100, 2, 3, 7, 100, 6, 0, \dots) \\ N_P &= 2^4 3^{100} 5^5 6^{100} 7^2 11^3 13^7 17^{100} 19^6. \end{aligned}$$

Since we know P is a consequence of the integers, it follows that $g(N_P) = 1$. Notice there will be many integers that do not even correspond to sentences and those are sent to 0 as well.

We have self-encoded logic in a single model of the integers. Therefore what Gödel did and what we will do is simply ask our lie-detector if every consequence has a proof. We do this by creating the sentence

$$Q(P) \equiv \text{"The sentence } P \text{ cannot be proved."}$$

This is a sentence Q with a variable P and so it has no intrinsic value of true or false so it is not possible to ask if $Q(P)$ is a consequence. Only after substituting for P can we reach the claim that Q is true of false. However, every sentence P can be converted into an integer, as we did above. Hence, instead of testing if $Q(P)$ is a consequence for some substitution P, we create a function $h : \mathbb{N} \to \{0, 1\}$ as follows. If $n = N_P$ for a sentence P, then $h(n) = g(n_{Q(P)})$, and 0 otherwise. That is, h outputs 1 only if we substitute in a value n corresponding to a sentence P which makes $Q(P)$ a consequence for the integers. We ask, is $h(n_Q) = 0$ or is $h(n_Q) = 1$? The reason to ask that question is because of the paradox that ensues.

$$Q(Q) \equiv \text{"The sentence } Q \text{ cannot be proved."}$$

So, if $h(n_Q) = 1$ then $Q(Q)$ is a consequence, i.e. it is true. Since $Q(Q)$ is true, it means that Q is a sentence that cannot be proved (yet Q is true). On the other-hand, if $h(n_Q) = 0$ then $Q(Q)$ is not a consequence, i.e. it is false. But $Q(Q)$ is false is to say: the sentence Q can be proved (yet Q is false)!

Most mathematicians and philosophers take the view that the integers are self-evident and so there cannot be proofs to falsehoods. Hence, we are accustomed to viewing Gödel's theorem as stating that there are truths about integers which cannot be proved by properties of the integers. The next question becomes:

> Can we know when a sentence is a consequences without a proof?

Turing answered no! The process is known as *Turing's Halting Problem*. It works as follows. Suppose that we give a computer a program P and initial values $a_1, \ldots, a_n$. The computer runs the program with the initial settings $a_1, \ldots, a_n$, but doing so it may run forever or it may stop after some amount of time. Turing asked,

> Is there a program Q that given as input $[P, a_1, \ldots, a_n]$, determines if P will run forever on the inputs $a_1, \ldots, a_n$?

Notice Q cannot simply run the program P because P may never stop and that would prevent Q from correctly answering that P will not halt. So Q must some how determine if the logic inside of P is so designed as to avoid running forever for the given input but without actually running through the logic step by step. Said another way, Q will prove that $[P, a_1, \ldots, a_n]$ halts without *giving the steps to prove it halts* because those steps are the actually running of the program P. If Turing's question is answered as false, then we there are consequences which a computer cannot predict are consequences. Turing demonstrated that if there is a program Q it cannot decide if Q itself will halt. Therefore, Q cannot exist (it is its own counter-example).

Remark 1.7.2. We say a theory T for the axioms Φ is *complete* if every consequence can be reached by a proof starting from the axioms Φ. We also say that T is *consistent* if every proof reaches a consequence of the axioms. So Gödel's theorem can be stated as:

> The theory of the integers is consistent if, and only if, it is not complete.

Say a theory T is *decidable* if there is an algorithm to determine if a sentence is in the theory. So Turing's theorem can be expressed as:

> The integers are undecidable.

Chapter 2

Congruence, Quotients, and Epimorphisms

Motivation

Consider how we solve for an unknown value x in equations such as $x + 2 = 5$. If we are careful to record each of our steps we would likely proceed as follows:

$$\begin{aligned} x + 2 &= 5; \\ (x + 2) + (-2) &= 5 + (-2); \\ x + (2 + (-2)) &= 3; \\ x + 0 &= 3 \\ x &= 3. \end{aligned}$$

Each step above reveals something significant about our understanding of what we now collectively regard as *algebra*. In time we will unpack each of these steps and see how widely they apply. For this first chapter we want to focus on the first and perhaps most essential algebraic tool – the ability to apply operations to both sides of an equal sign and maintain an equality. For example, we did this when we wrote $(x+2)+(-2) = 5+(-2)$, the operation was addition. Now let us inspect a more subtle example. Suppose we solve for x in $x^2 = 25$. We might proceed as follows:

$$x^2 = 25; \qquad x = \pm\sqrt{25}; \qquad x = \pm 5.$$

The ability to write $x = \pm\sqrt{25}$ suggests cooperation between equality and the operation of $\pm\sqrt{}$. Unfortunately, this cooperation is fictitious as what we have actually done is split our reasoning into two separate cases: $x = \sqrt{25}$ or $x = -\sqrt{25}$. When our reasoning is forced to branch into cases, their is no predictable outcome as the answer depends on the choices we make. Of course, we recognize that 'the' solution is the union of all possible outcomes so in a meaningful way we want to treat $\sqrt{25}$ and $-\sqrt{25}$ as equivalent even though they are evidently different numbers. We cannot do this unless we broaden our notion of equality. So this is the route we take next.

First we describe a sensible generalization of equality known as *equivalence* and we also describe the relationship of this construction to several other important topics including partitions and functions. Then we return to algebra and describe what it means for an equivalence and an operation (say addition or the taking of roots) to coexist in a manner that permits us to solve equations by algebraic methods. That concept is known as *congruence*. Finally, we extend the purely set-theoretic relationships between equivalence, partitions, and functions to congruence, quotients, and homomorphisms. The result is what is commonly known as the *Fundamental Homomorphism Theorem*.

2.1 Equivalence

Further Reading: Jacobson §0.4

Throughout this section we temporarily ignore operations such as addition, multiplication, squaring, etc. and focus simply on sets and special relations on sets.

2.1.1 Equivalence relations

Recall that a relation on non-empty sets (or classes) S and T is a subset (or subclass) R of the set (or class) $S \times T$ of ordered pairs (s, t), for $s \in S$ and $t \in T$. If $(s, t) \in R$ we will write sRt. We now concentrate on a special relation which appears throughout Set Theory and mathematics.

Definition 2.1.1. A relation E on A (more precisely on $A \times A$) is an *equivalence* if it satisfies the following combined properties:

Reflexive for all $a \in A$, aEa;

Symmetric for all $a, b \in A$, if aEb then bEa; and

Transitive for all $a, b, c, \in A$, if aEb and bEc then aEc.

Notice that the usual equal sign is always an equivalence relation, regardless of the set in question (actually this is a somewhat interesting result for the inquisitive minded and so we include the proof as an exercise below). However, we are quite familiar and comfortable with relations that are not equivalences. As already described, $a \leq b$ is a relation on $\mathbb{Z}$ (or formally on $\mathbb{Z} \times \mathbb{Z}$); yet, $\leq$ is *not* an equivalence because it is not symmetric. To illustrate how we verify a relation is an equivalence we pick just one of many interesting examples and leave others to exercises.

Proposition 2.1.2. *Let $E = \{(a, b) \in \mathbb{R} \times \mathbb{R} : a - b$ is an interger multiple of $2\pi\}$. It follows that E is an equivalence relation on $\mathbb{R}$.*

Proof. First we show E is reflexive. For every $a \in \mathbb{R}$, $a - a = 0$ is an integer multiple of 2π, namely $0 \cdot 2\pi = 0 = a - a$. Therefore $(a, a) \in E$, or in the usual infix notation aEa.

Secondly we show E is symmetric. Thus we suppose there are $a, b \in \mathbb{R}$ such that aEb. This means that $a - b$ is an integer multiple of 2π, that is, that $a - b = 2\pi k$ for some integer k. Thus, $b - a = -2\pi k = 2\pi(-k)$. Since $-k$ is an integer, we have shown that $b - a$ is an integer multiple of 2π. Hence, bEa.

Finally we show E is transitive. First suppose there are $a, b, c \in \mathbb{R}$ such that aEb and bEc. Hence there are integers k and j such that $a - b = 2\pi k$ and $b - c = 2\pi j$. Therefore,

$$a - c = (a - b) + (b - c) = 2\pi k + 2\pi j = 2\pi(k + j).$$

As $k + j$ is an integer we conclude that aEc.

As E is reflexive, symmetric, and transitive, it is an equivalence relation. □

2.1 Modulo 12 Suppose that $E_{12} = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : a - b$ is a multiple of $12\}$. Prove that E_{12} is an equivalence relation on $\mathbb{Z}$. Also, replace 12 with an arbitrary integer n and so E_n remains an equivalence relation on $\mathbb{Z}$.

2.2 True or False? If $E = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : a + b$ is a multiple of $12\}$, is E and equivalence relation on $\mathbb{Z}$?

2.3 Fraction equality Let $Q = \mathbb{Z} \times \mathbb{Z}^+$ (where $\mathbb{Z}^+ = \{1, 2, \dots\}$). Suppose that $E = \{((a, b), (c, d)) \in Q \times Q : ad = bc\}$. Prove that E is and equivalence on Q.

Moving forward we will now write equivalence relations with suggestive symbols such as $\equiv$, $\cong$, $\sim$, etc. For example, we might write $x + 2 \equiv 5$ which will mean to consider $x + 2$ as equivalent to 5 but only to within the notion of equality given to us by $\equiv$. For instance, if we use the relation E_{12} of Exercise 1 then $x + 2 \equiv 5$ will mean simply that $(x + 2) - 5$ is a multiple of 12. Because such equivalences are vastly important they receive special notation. We write:

$$a \equiv b \pmod{n} \text{ if, and only if, } a - b \text{ is a multiple of } n. \tag{2.1}$$

Now we briefly detour back to Set Theory to introduce the formal meaning of the symbol $=$. Recall, in our understanding every*thing* is a set. So it may appear

that we have already been using = without understanding it. This is true of our examples but not of our definition of equivalence relations and so we remain on logically stable ground. Arguably the following exercise should precede all uses of =, but recall our goal is not build from perfect axioms upward but to start from what we already know and build outward in all directions, including foundations.

2.4 Set equalaity Two sets A and B are called *equal* if A is a subset of B and B is a subset of A. In symbols we write $A = B$ if, and only if, $A \subseteq B$ and $B \subseteq A$. Show that = is an equivalence relation for the class of all sets.

We have defined sets as any collection of objects, substance, etc. This makes one family of sets stand out, the empty collections, or rather the *empty sets*. The physically minded may have many examples of such objects. For instance quarks, electrons, neutrinos, etc. Any object that is not comprised of subobjects could fairly be described as an "empty set". However, the definition of equal sets given in Exercise 2.1.1 is a bit narrow minded when it comes to such distinctions. For as we see in Exercise 2.1.1, below, if that is the meaning of equivalent sets, then two empty sets are equal. So Set Theory sees no difference between quarks and electrons, for example. Thus, the mathematics of Set Theory is not intending to describe the universe as we might encounter it within physics.[1]

2.5 Unique emptyset Under the definition of set equivalence in Exercise A, show that two sets that have no subsets must be equal. That is, to Set Theory (and to mathematics in general) there is only *one* empty set and we denote it by $\varnothing$.

2.6 Show that under the definition of equality in Exercise 2.1.1, it follows that $\{3, 2, 1\} = \{1, 2, 2, 3, 1\}$. In particular observe that sets do not record information about position of an object in a list nor do they record information about the number of times an object appears in a list.

Remark 2.1.3. Exercise 2.1.1 indicates an important option for computer programs that implement sets as a data structure. In most computer systems data is stored implicitly or explicitly with information including location and the number of times the same item appears. However, there are many uses for the mathematical construction of a set. The problem is that simple tasks, such as adding an element to a set, become laborious if not thought out properly. For instance, if the computer has stored a set as a list $\{1, 2, 3, \ldots, 1000\}$ and the user adds 99, the algorithm may be forced to search through the list to discover that 99 already exists in the set and therefore need not be added. When a set will be updated many times in a row this can be a very slow process. However, the definition of sets does not insist that no duplicates occur. Instead, it only requires that if another set has the same entries but in different quantities, then these two sets should be considered as equal. So it is essential only to implement an equality test for sets and the actual data can indeed be stored in any order and with any number of repeats.

2.1.2 Partitions

So far we have viewed equivalence as a relationship between elements in a set. However, it is often helpful to visualize these relationships. One device well suited for that task is the notion of a partition of a set.

[1] There are two possible adjustments here. We could argue the quarks, electrons, etc. are not empty because they consist of values such as the flavor, charge, and mass. Still another alternative is to argue with strings – this is perhaps even further removed from Set Theory foundations.

Definition 2.1.4. A *partition* on a set A is a set $\mathscr{P}$ of nonempty subsets of A such that

(i) for every $a \in A$ there is a $P \in \mathscr{P}$ such that $a \in P$, and

(ii) for every $P, Q \in \mathscr{P}$, if $P \cap Q \neq \varnothing$ then $P = Q$.

There are various other ways to express the definition of a partition. For instance, $\mathscr{P}$ is a partition if for every $a \in A$ there is a unique $P \in \mathscr{P}$ such that $a \in P$. Though that definition is certainly more compressed, conversely it often leads to longer proofs. Indeed, the shorter the definition the more likely a subtle property has been hidden behind some layers of reasoning that must then become part of every proof involving the definition. For this reason, it is a good policy to settle on definitions that balance the need to have compressed memorable properties with the ease of executing proofs.

Proposition 2.1.5. *Fix an integer m and define for each integer n, the set*

$$n + m\mathbb{Z} = \{n + ms : s \in \mathbb{Z}\}.$$

Define $\mathscr{P}_m = \{n + m\mathbb{Z} : n \in \mathbb{Z}\}$. It follows that $\mathscr{P}_m$ is a partition of $\mathbb{Z}$.

Proof. First, for every integer n, $n = n + m \cdot 0 \in n + m\mathbb{Z}$ and so $\mathscr{P}_m$ is indeed a set of nonempty subsets of $\mathbb{Z}$. Furthermore, this also shows that every integer is contained in one of the members of $\mathscr{P}_m$. So toFix an equivalence relation E on a set X and a partition $\mathscr{P}$ on X. The following hold: demonstrate that $\mathscr{P}_m$ is a partition it remains simply to show that if $n + m\mathbb{Z}$ intersects $n' + m\mathbb{Z}$ nontrivially, then $n + m\mathbb{Z} = n' + m\mathbb{Z}$. (Note that we are *not* suggesting that $n = n'$.)

So we begin by taking k in the intersection of $n + m\mathbb{Z}$ and $n' + m\mathbb{Z}$. We start by showing that $n + m\mathbb{Z}$ is a subset of $n' + m\mathbb{Z}$. For each $x \in n + m\mathbb{Z}$, $x = n + ms$ for some $s \in \mathbb{Z}$. Because of where we found k, we also know that for some integers t and u, $k = n + mt$ and $k = n' + mu$. That is $n = k - mt$ and $n' = k - mu$. Thus,

$$\begin{aligned} x = n + ms &= (n' - n') + n + ms \\ &= n' - (k - mu) + (k - mt) + ms \\ &= n' + mu - mt + ms = n' + m(s - t + u). \end{aligned}$$

So we have shown that $x \in n' + m\mathbb{Z}$ and so $n + m\mathbb{Z} \subseteq n' + m\mathbb{Z}$. Now swapping the roles of n and n' we see that also $n' + m\mathbb{Z} \subseteq n + m\mathbb{Z}$ and so indeed $n + m\mathbb{Z} = n' + m\mathbb{Z}$.[2]

This proves that $\mathbb{P}_m$ is a partition of $\mathbb{Z}$. □

As we claimed, there is a way to visualize partitions. For example we can

[2] In many situations arguments that can be repeated by simply replacing one variable with another are not worth re-writing. The Latin phrase *mutatis-mutandis* is often used to express this to a reader in an intimidating way. In truth, elegant writing, such as that of Halmos, can achieve a brief proof without resorting to this trick. Nevertheless, this is common.

visual the partition $\mathscr{P}_3$ of $\mathbb{Z}$ as follows.

$\mathbb{Z}$	$\cdots$	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	$\cdots$
$\vdots$														
$-2+3\mathbb{Z}$			-5			-2			1			4		
$-1+3\mathbb{Z}$				-4			-1			2			5	
$0+3\mathbb{Z}$		-6			-3			0			3			
$1+3\mathbb{Z}$			-5			-2			1			4		
$2+3\mathbb{Z}$				-4			-1			2			5	
$3+3\mathbb{Z}$		-6			-3			0			3			
$4+3\mathbb{Z}$			-5			-2			1			4		
$\vdots$														

Notice we have singled out with horizontal bars the three sets $0+3\mathbb{Z}$, $1+3\mathbb{Z}$, and $2+3\mathbb{Z}$. These make up the partition. All others $n+3\mathbb{Z}$ are simply other names for one of these three sets, for example, notice the rows for $-1+3\mathbb{Z}$ and $2+3\mathbb{Z}$ are identical.

The partition we introduced in Proposition-2.1.5 is incredibly useful and has its own important notation. From now on we will always write this partition as

$$\mathbb{Z}/m\mathbb{Z} = \{n + m\mathbb{Z} : n \in \mathbb{Z}\}. \tag{2.2}$$

In some circles this is also denoted by $\mathbb{Z}/m$ or even $\mathbb{Z}_m$.

2.7 True or False? The set $\{1 + m\mathbb{Z} : m \in \mathbb{Z}\}$ is a partition of $\mathbb{Z}$.

2.8 For every real number x let

$$x + 2\pi\mathbb{Z} = \{x + 2\pi k : k \in \mathbb{Z}\}.$$

Set $\mathbb{R}/2\pi\mathbb{Z} = \{x + 2\pi\mathbb{Z} : x \in \mathbb{R}\}$. Show that $\mathbb{R}/2\pi\mathbb{Z}$ is a partition of $\mathbb{R}$.

In Figure 2.1 we give a visual description of the partition we created in Exercise 5. Notice the real number line is bent into a spiral to create the geometric correspondence, and the vertical lines demonstrate points that are equivalent under the definition given in Exercise 5.

Partitions are wonderful topic in their own right but our interests require that we connect them also to external concepts and we do that first with equivalence relations.

If E is an equivalence relation on a set X then define for each $x \in X$ the *equivalence class* of x, namely

$$[x] := \{y \in X : xEy\}. \tag{2.3}$$

$$X/E = \{[x] : x \in X\}. \tag{2.4}$$

If $\mathscr{P}$ is a partition on X, then for $x, y \in X$, define

$$x \equiv_{\mathscr{P}} y \text{ if, and only if, } \exists P \in \mathscr{P}, x, y \in P. \tag{2.5}$$

With that notation we prove that equivalence relations are essentially the same as partitions.

Proposition 2.1.6. *Fix an equivalence relation E on a set X and a partition $\mathscr{P}$ on X. The following hold:*

(i) X/E is a partition on X,

(ii) $\equiv_{\mathscr{P}}$ is an equivalence relation on X, and

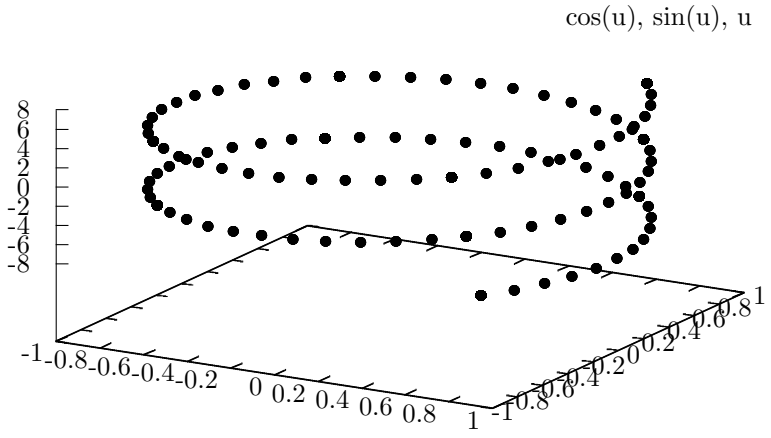


Figure 2.1: The helix is the natural way to describe the partition $\mathbb{R}/2\pi\mathbb{Z}$. Notice the real line can be bent into the helix shape which winds above the unit circle. Two points on the helix lie in the same member of the partition if, and only if they lie, on the same vertical line – which we often call "fibers". So in one revolution along the edge of the helix we visit every member of the partition exactly once. In two revolutions we visit each member twice.

(iii) E is the same equivalence ralation as $\equiv_{X/E}$, i.e. aEb if, and only if, $a \equiv_{X/E} b$. Also, the partition $\mathscr{P}$ is the same as $X/(\equiv_{\mathscr{P}})$, i.e. $P \in \mathscr{P}$ if, and only if, $P \in X/(\equiv_{\mathscr{P}})$.

Proof. For (i), observe that for every $x \in X$, $[x]$ is nonempty as E is reflexive. Furthermore, every $x \in X$ lies in $[x]$ so the first property of a partition is satisfied.. Also, if $[x] \cap [x'] \neq \varnothing$ then there is a $y \in X$ such that xEy and $x'Ey$. Therefore xEy and yEx' so by the transitive property xEx'. Now if for all $z \in [x']$, we have xEx' and $x'Ez$ so again by the transitive property xEz. Therefore $z \in [x]$ so that $[x'] \subseteq [x]$. By swapping the roles of x and x' we find also $[x] \subseteq [x']$. Thus $[x] = [x']$. Thus, the second property of a partition has been proved. So X/E is a partition.

For (ii), let $x \in X$. There is a unique $P \in \mathscr{P}$ such that $x \in P$ and so $x, x \in P$ which means that $x \equiv_{\mathscr{P}} x$, i.e. the reflexive property holds. Next if $x, y \in X$ and $x \equiv_{\mathscr{P}} y$ then there is a $P \in \mathscr{P}$ such that $x, y \in P$. Thus $y, x \in P$ and so $y \equiv_{\mathscr{P}} x$ – the symmetric property. Finally, if $x, y, z \in X$ such that $x \equiv_{\mathscr{P}} y$ and $y \equiv_{\mathscr{P}} z$ then there are $P, Q \in \mathscr{P}$ such that $x, y \in P$ and $y, z \in Q$. However this implies that $y \in P \cap Q$ so $P \cap Q \neq \varnothing$. Thus, $P = Q$ and so $x, z \in P$ proving that $x \equiv_{\mathscr{P}} z$ – the transitive property.

Part (iii) is crucial but we leave it as an exercise which will help practice the meaning of the notation. □

2.9 Complete the proof of Proposition 2.1.6(iii).

2.10 Consider the relation $\leq$ on $\mathbb{R}$. For $x \in \mathbb{R}$, define $[x] = \{y \in \mathbb{R} : x \leq y\}$ and $\mathbb{R}/_{\leq} = \{[x] : x \in \mathbb{R}\}$. This mimics what we defined above however, show that $\mathbb{R}/_{\leq}$ is not a partition. Also show that $\leq$ is not an equivalence relation and so this is not a contradiction of Proposition 2.1.6(i).

2.11 A function $f : \mathbb{R} \to \mathbb{R}$ is *even* if for all $x \in \mathbb{R}$, $f(-x) = f(x)$, *odd* if for all $x \in \mathbb{R}$, $f(-x) = -f(x)$.

(i) Show that some functions $f : \mathbb{R} \to \mathbb{R}$ are neither even nor odd.

(ii) Let S be the set of all functions $f : \mathbb{R} \to \mathbb{R}$ which either even or odd. Decide if the $\{f \in S : f \text{ is even }\}$ and $\{f \in S : f \text{ is odd }\}$ partition S. [Hint: are some functions even and odd?]

2.1.3 Functions and partial functions

Further Reading: Jacobson §0.3

We have discovered sets and classes allow us to formalize many concepts once known intuitively. So far we have worked with sets by relations and partitions. The third powerful tool is the concept of a function. Indeed, functions are perhaps the best known mathematical object after numbers and so we hope that our following technical description will not supplant the intuition gained by years of experience with functions. However, as we move our reasonnig increasingly further away from real numbers and elementary functions that we can graph, it becomes necessary to declare the formal expectations about functions.

The most familiar definition of a function is a relation from a set A to a set B for which every $a \in A$ is related to one and only one $b \in B$. We want also the intermediate concept of a partial function.

Definition 2.1.7. A *partial function* $f : A \to B$ between sets (or classes) A and B is a relation $R \subseteq A \times B$ such that:

if for some $a \in A$ and $b, b' \in B$, aRb and aRb', then $b = b'$.

We say that R is *well-defined* when that property holds and we write $f(a) = b$ since no ambiguity can occur about which element of B to associate to a.

- The *domain*, denoted $\operatorname{dom} f$, of a partial function is $\{a : (a, b) \in R\}$.
- The *image*, denoted $\operatorname{im} f$, of a partial function is $\{b : (a, b) \in R\}$.
- The *codomain*, denoted $\operatorname{codom} f$, of a partial function is B.

Finally, a *function* is a partial function whose domain is A.

Our assumptions allow us to define *fibers* for partial function $f : A \to B$. For each $b \in B$ define

$$f^{-1}(b) = \{a \in \operatorname{dom} f : f(a) = b\}. \tag{2.6}$$

Definition 2.1.8. A *surjection* is a function $f : X \to Y$ such that for each $y \in Y$ there is an $x \in X$ such that $f(x) = y$. We also say that f is *surjective* or that it is *onto*.

We have seen many examples of surjections already, as the following result makes clear.

Proposition 2.1.9. *Fix a set X.*

(i) If $f : X \to Y$ is surjective function then

$$X/f = \{f^{-1}(y) : y \in Y\}$$

is a partition of X.

(ii) If $\mathscr{P}$ is a partition of X then define $f_{\mathscr{P}} : X \to \mathscr{P}$ as follows: $f_{\mathscr{P}}(x) = P \in \mathscr{P}$ if, and only if, $x \in P$. It follows that $f_{\mathscr{P}}$ is a surjection and $\mathscr{P} = X/f_{\mathscr{P}}$.

Proof. Let $f : X \to Y$ be a surjection. For every $y \in Y$ there is an $x \in X$ such that $f(x) = y$. Hence, $x \in f^{-1}(y)$ and so $f^{-1}(y) \neq \varnothing$.

Next, f is a function (not a partial function) and so for every $x \in X$, there is a value $f(x) = y$. Hence, $f^{-1}(y)$ contains x and so $x \in f^{-1}(y) = f^{-1}(f(x))$.

Finally, suppose that for some y and y' in Y, $f^{-1}(y) \cap f^{-1}(y') \neq \varnothing$. Thus, there is an $x \in f^{-1}(y)$ and $x \in f^{-1}(y')$. This means that $y = f(x) = y'$ and so $y = y'$. Therefore, $f^{-1}(y) = f^{-1}(y')$. We conclude that X/f is a partition. This proves (i).

Part (ii) is left as an exercise. □

2.12 Complete the proof of Proposition-2.1.9. In particular show that $f_{\mathscr{P}}$ is a function (i.e. that its domain is X and that is is well-defined). Then show $f_{\mathscr{P}}$ is surjective. Finally show that the set of fibers of $f_{\mathscr{P}}$ is the set $\mathscr{P}$.

2.13 Determine a surjection from $\mathbb{R}$ such that the induced partition $\mathbb{R}/2\pi\mathbb{Z}$ (as defined in Exercise 2.1.2).

Now, the precise correspondence is between equivalence, partitions, and surjective functions. However, it is not too difficult to relax the surjective assumption to general functions simply by observing that if $f : A \to B$ is not surjective, then we can form a new function $f' : A \to \operatorname{im} f$ where $f'(a) = f(a)$, for each $a \in A$. Notice that f' is surjective even if f was not.

We do a similar thing to turn partial functions $g : A \to B$ into functions. There we create a function $g^* : \operatorname{dom} g \to B$ such that $g^*(a) = g(a)$, for each $a \in \operatorname{dom} g$. Of course, g^* is now a function and its very definition is gleaned from g but simply adapts the input to come from the domain.

Now because these two associated functions f' and g^* are so naturally related to the original f and g, for the most part we do not bother to issue them special notation. So even though we used f' and g^* above, we will not repeat this and instead we will write simply $f : A \to \operatorname{im} f$ or $g : \operatorname{dom} g \to B$, unless we detect that this will produce a problem.

Finally, to every partition $A/f = \{f^{-1}(b) : b \in B\}$ into fibers of a surjective function $f : A \to B$ we can create a new function, $\hat{f} : A/f \to B$ as follows:

$$\hat{f}(f^{-1}(b)) = b \qquad (\forall b \in B). \tag{2.7}$$

This may seem somewhat silly but we have seen that often silly notions (such as the reflexive property) have there uses. The importance of $\hat{f}$ is that it shows us one way to rewrite a function as a composition of two. If we let $\check{f} : A \to A/f$ be defined as

$$\check{f}(a) = f^{-1}(f(a)) \qquad (\forall a \in A) \tag{2.8}$$

then we find that $f = \hat{f}\check{f}$ (or rather $f = \hat{f} \circ \check{f}$ in classic composition notation). Notice that $\check{f}$ is surjective where as $\hat{f}$ is invertible. In a vague sense (which Exercise 3 below will tighten up) the two functions f and $\check{f}$ are essentailly the same. This allows us to conclude by noticing that to induce a partition on a set A we can specify many many possible surjections for the same partition (Exercise 4). Therefore, although it may seem that we have said partitions are the same as surjections, this is only true if equate all surjections in the manner desribed in Exercise 3. In general, there are many more surjections from A than possible partitions on A.

2.14 Define a relation between functions as follows. Say that $f : A \to B$ is equivalent to $g : A \to C$ if there is an invertible function $h : B \to C$ such that $hf = g$ (i.e. that $h \circ f = g$). We write $f \sim g$. Show that $\sim$ is an equivalence relation on the class of all functions.

2.15 Under the equivalence $\sim$ above, show that if $f : A \to B$ and $g : A \to C$ such that $f \sim g$ then $A/f = A/g$.

2.16 How many partitions are there of $\{1, 2, 3\}$? How many surjections are there with $\{1, 2, 3\}$ as the domain? Of the possible surjections from $\{1, 2, 3\}$, many different equivalence classes are there with respect to the equivalence $\sim$ of above?

2.2 Congruence

Further Reading: Jacobson §1.1, §1.8

Here we begin the study of equivalence as it appears in algebra. This is similar in many ways to equivalence amongst sets except that we now must take care to consider operations on the sets.

Recall that an operation on a set S is a function on several copies of S into a single copy of S. For example addition of integers $a + b$ is a function on pairs (a, b) to a single integer $a+b$ so the function's domain is $\mathbb{Z} \times \mathbb{Z}$ and its codomain is $\mathbb{Z}$. Multiplication $\cdot$ also maps $\mathbb{Z} \times \mathbb{Z}$ into $\mathbb{Z}$. These are the most common and we call them *binary operations*. Negatives however operate on only one copy of $\mathbb{Z}$, e.g. -5. So negatives are a function $\mathbb{Z} \to \mathbb{Z}$. Those we call *unary operations*. The made-up operation $[x, y, z] = x^2 + yz$ would be treated as an operation from $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ into $\mathbb{Z}$. If that needs a name it would be called a *ternary operation* or a simply an operation with *signature 3*. Constants such as 0 and 1 can be also be viewed as operations that require no input, but take a single value in $\mathbb{Z}$. So these have signature 0 and are sometimes called *nullary operations*. Just to get the general ideas out we present the definitions and results here in terms of any operation. So we write this as $[s_1, \ldots, s_n]$ similar to our made up example with 3 variables. But it will be immediately important to try the meaning out with familiar examples such as $+$, $\cdot$, $-$, 0, and 1.

Definition 2.2.1. A *congruence* for an operation $[s_1, \ldots, s_n]$ on a set S is an equivalence relation $\equiv$ on S such that

$$\text{if } s_1 \equiv t_1, \ldots, s_n \equiv t_n \text{ then } [s_1, \ldots, s_n] \equiv [t_1, \ldots, t_n].$$

For example, the usual equality of decimal numbers in $\mathbb{R}$ is a congruence for the (binary) operation of addition. That gives us the right to write:

$$\text{As } x + 2 = 5 \text{ it follows that } (x + 2) + (-2) = 5 + (-2).$$

It may not appear that we have just involved a congruence but we have. To see it, first suppose we make the awkward move to write $[a, b]$ for $a + b$ and $\equiv$ for equality of decimal numbers. This is not necessary but it helps us better spot what is happening. For example, the seemingly pointless observation that $-2 = -2$ is now written as $-2 \equiv -2$, and because the symbols are new it suggest we be a little more careful in trusting this claim. In this way we could have written our sentence as:

$$([x, 2] \equiv 5) \text{ and } (-2 \equiv 2) \text{ which by congruence rules implies } [[x, 2], -2] \equiv [5, -2].$$

This now reflects our formal definition of a congruence. As we should expect, that notation is correct but tedious. In general it is perfectly sensible to continue in the short-hand natural notations we have practiced for so long with our real

numbers. We simply include this translation as an example of how to read and use the definition of congruence.

2.17 Define $(a,b) \equiv (c,d)$ by $ad = bc$, for $(a,b),(c,d) \in \mathbb{Z} \times \mathbb{Z}^+$. Also define a binary operation $(a,b) + (c,d) = (ad + bc, bd)$. Show that $\equiv$ is a congruence for $+$. Notice, if write a/b instead of (a,b) then we have just described the usual meaning of equality of fractions and addition of fractions.

2.18 Define $(a,b) \oplus (c,d) = (a + c, b + d)$ for $(a,b),(c,d) \in \mathbb{Z} \times \mathbb{Z}^+$.

(i) Decide if $=$ is a congruence for $\oplus$ (where $(x,y) = (z,w)$ means that $x = z$ and $y = w$).

(ii) Decide if $\equiv$ (from Exercise 1 above) is a congruence for $\oplus$.

The equality of decimal numbers is quite a powerful congruence as it is a congruence for the unary operation of inversion, e.g. $x = 5$ so $x^{-1} = 5^{-1}$, and the binary operations of addition, subtraction, multiplication, division, and exponentiation. It may seem therefore that we should generalize the meaning of congruence to many operations. While this is certainly possible it is also not necessary if we bother to be modestly clever. For example, instead of thinking of $\mathbb{R}$ as having two binary operations, addition and multiplication, we might consider it has having solely one *ternary* operation $\ell : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

$$\ell(m,x,b) = m \cdot x + b.$$

The letter ℓ should remind the reader of 'line' which is evidently the structure underpinning this ternary product. Notice that we can recover addition from ℓ by simply looking at $\ell(1,a,b) = 1 \cdot a + b = a + b$ and we can also recover multiplication by using $\ell(a,b,0) = a \cdot b + 0 = a \cdot b$. What has happened is that we have now forever linked the very geometric idea of a line with the very algebraic idea of addition and multiplication. So, if we want to study decimal equality as a congruence for both addition and multiplication we are actually studying equality of lines. So we have a choice to either study $\langle \mathbb{R}, +, \cdot \rangle$ with two binary operations or one ternary operation ℓ. It is simply a matter of taste but for brevity we will elect to study congruences with respect to a single operation at a time, allowing that operation to be ternary or n-ary if we like. For further reading on the ternary operation of forming lines consider M. Hall *The Theory of Groups*.

One last comment on operations is that they often transfer from one set to another. For example, let X be a set. Define $\mathrm{fun}(X,\mathbb{R})$ as the set of all functions $f : X \to \mathbb{R}$. We instantly have a notion of addition in $\mathrm{fun}(X,\mathbb{R})$, namely, $f + g : X \to \mathbb{R}$ is defined as $(f+g)(x) = f(x) + g(x)$. We can also define $f \cdot g : X \to \mathbb{R}$ so that $(f \cdot g)(x) = f(x)g(x)$. We call these natural operations *pointwise* or *componentwise* operations.

2.2.1 Quotients

Our next consideration is the matter of replacing equivalence with partitions and then also surjections. We saw that those objects offer an alternative and robust translation of the notion of equivalence and we should like to use this together with algebra and congruence.

Definition 2.2.2. Let $\langle S, [x_1, \ldots, x_n] \rangle$ be a set with an n-ary operation. A *quotient* $\mathscr{Q}$ of S with respect to $[x_1, \ldots, x_n]$ is a partition $\mathscr{Q}$ of S such that for all $P_1, \ldots, P_n \in \mathscr{Q}$, and all $s_i \in P_i$, there exists a unique $Q \in \mathscr{Q}$ such that $[s_1, \ldots, s_n] \in Q$. We denote Q as $[P_1, \ldots, P_n]$.

When a partition is a quotient of a set with an operation we sometimes say that the partition "admits the operation". Whenever the operation in question is understood from the context we permit ourselves to speak of the "a quotient of S" rather than "...a quotient of S with respect to...".

Proposition 2.2.3. *For every integer m, $\mathbb{Z}/m\mathbb{Z}$ is a quotient of $\mathbb{Z}$ with respect to addition.*

Proof. Recall members of $\mathbb{Z}/m\mathbb{Z}$ take the form $x + m\mathbb{Z}$ for an integer $x \in \mathbb{Z}$. We must show that for all $x + m\mathbb{Z}, y + m\mathbb{Z} \in \mathbb{Z}/m\mathbb{Z}$, there exists a unique $z + m\mathbb{Z} \in \mathbb{Z}/m\mathbb{Z}$ such that for all $x + ms \in x + m\mathbb{Z}$ and all $y + mt \in y + m\mathbb{Z}$, $(x + ms) + (y + mt) \in z + m\mathbb{Z}$. Since $z + m\mathbb{Z}$ is unique to $x + m\mathbb{Z}$ and $y + m\mathbb{Z}$ we can discover the correct z by choosing to consider just x and y (i.e. letting $s = 0$ and $t = 0$).[3] Certainly, $x+y = (x+y)+m\cdot 0$ and so $x+y \in (x+y)+m\mathbb{Z}$. Therefore, if an appropriate z exists, then we might as well assume $z = x + y$. Now we show that such a choice of $z = x + y$ is sufficient.

For all s and t, $(x + ms) + (y + mt) = (x + y) + m(s + t) \in (x + y) + m\mathbb{Z}$. Therefore, $\mathbb{Z}/m\mathbb{Z}$ admits $+$ and so it is a quotient of $\mathbb{Z}$ with respect to addition. □

2.19 Show that $\mathbb{R}/2\pi\mathbb{Z}$ is a quotient of $\mathbb{R}$ with respect to addition. Is a quotient of $\mathbb{R}$ with respect to a multiplication?

2.20 Show that for each integer n, $\mathbb{Z}/n\mathbb{Z}$ is a quotient of $\langle \mathbb{Z}, \cdot \rangle$.

2.21 Let $S = \mathbb{R} \times \mathbb{R}$ be the xy-plane with the usual addition of vectors, i.e. $(a, b) + (x, y) = (a + x, b + y)$. Let $\mathscr{Q}$ be the set of all lines parallel to $y = 2x$. Show that $\mathscr{Q}$ is a quotient of $\langle S, + \rangle$.

Having tried some examples we are prepared to relate congruence to quotients in the following formal result.

Proposition 2.2.4. *Let $\equiv$ be an equivalence relation on a set S with associated partition $\mathscr{P} = S/_{\equiv}$. It follows that $\equiv$ is a congruence for an operation $\{s_1, \ldots, s_n\}$ on S if, and only if, $S/_{\equiv}$ is a quotient of S with respect to $[s_1, \ldots, s_n]$.*

Proof. Let $\equiv$ be a congruence for an operation $[x_1, \ldots, x_n]$ on S. Take P_1 through P_n in $S/_{\equiv}$ and also take for all $i \in \{1, \ldots, n\}$ arbitrary $s_i, t_i \in P_i$. Note that for each i, $s_i \equiv t_i$ (by the definition of $S/_{\equiv}$). Thus, by the definition of congruence it follows that $[s_1, \ldots, s_n] \equiv [t_1, \ldots, t_n]$. In particular, if Q is the unique member of $S/_{\equiv}$ which contains $\{s_1, \ldots, s_n\}$ then Q also contains $[t_1, \ldots, t_n]$. In particular, $S/_{\equiv}$ admits the operation $[x_1, \ldots, x_n]$ on S and $Q = [P_1, \ldots, P_n]$. That is $S/_{\equiv}$ is a quotient.

Now suppose instead that a partition $\mathscr{P}$ admits an operation $[x_1, \ldots, x_n]$ and that $\equiv$ is the equivalence relation on S induced by $\mathscr{P}$. We must show $\equiv$ is a congruence for the operation. So take, for each $i \in \{1, \ldots, n\}$, $s_i, t_i \in S$ such that $s_i \equiv t_i$. These means that each s_i and t_i lie in the same member P_i of the partition $\mathscr{P}$. As $\mathscr{P}$ admits $\{x_1, \ldots, x_n\}$ it follows that $[s_1, \ldots, s_n], [t_1, \ldots, t_n] \in [P_1, \ldots, P_n]$ and so $[s_1, \ldots, s_n] \equiv [t_1, \ldots, t_n]$. So $\equiv$ is a congruence for $[x_1, \ldots, x_n]$. □

2.2.2 Homomorphisms

Further Reading: Jacobson §1.9

We are now finally prepared to involve functions in our study of operations. We should expect that the important functions will be surjections, because

[3] This is known as the "needle-in-the-haystack" heuristic. It says, if you search for something to exist uniquely, first consider the uniqueness, then you know what the element looks like and so the existence is probably obvious at that point.

these relate to equivalence relations. However, we know that not all equivalence relations are congruences (with respect to a fixed operation). So we should not expect that all surjections are equally useful and so we will need to introduce a subclass of functions that have just the properties we need to create congruences in a manner similar to how surjections produce equivalence relations. So we begin by inspecting that process.

Let $f : A \to B$ be a function. As we saw before, we can produce from f a natural surjective function which we still denote by f, namely $f : A \to \operatorname{im} f$. We therefore induce an equivalence relation on A by writing $a \equiv a'$ if, and only if, $f(a) = f(a')$. Now suppose that $\equiv$ is a congruence for a binary operation $[a, a']$ on A. This will mean that whenever $x \equiv x'$ and $y \equiv y'$ then $[x, y] \equiv [x', y']$. When we write this out with the function notation we are saying:

$$\text{if } f(x) = f(x') \text{ and } f(y) = f(y') \text{ then } f([x, y]) = f([x', y']).$$

This property is somewhat unhelpful because it places the equality between outputs of f and we have no operation on B which would explain a process to combine the items in the hypothesis to arrive at the conclusion. In fact, we should have no expectation that a function on the set A will have anything to do with the operation on A unless the function is somehow relating that operation to *another* operation.

So now let us reconsider the problem with some richer information.

Definition 2.2.5. A *homomorphism* from a set S with an n-ary operation $[s_1, \ldots, s_n]$ to a set T with an n-ary operation $[t_1, \ldots, t_n]$, is a function $f : S \to T$ with the property that

$$f([s_1, \ldots, s_n]) = [f(s_1), \ldots, f(s_n)] \qquad (\forall s_1, \ldots, s_n \in S).$$

We sometimes denote homomorphisms by $f : \langle S, [\ldots] \rangle \to \langle T, [\ldots] \rangle$.

2.22 Let S be the set of functions $f : \mathbb{R} \to \mathbb{R}$ that are even or ordd but non-zero (i.e. there is some $x \in \mathbb{R}$ with $f(x) \neq 0$). Define $\chi : S \to \mathbb{Z}/2\mathbb{Z}$ so that $\chi(f) \equiv 0 \pmod 2$ if f is even, $\chi(f) \equiv 1$ if f is odd.

(i) Prove that χ is a homomorphism of $\langle S, \cdot \rangle$ (even/odd functions under pointwise addition) to $\langle \mathbb{Z}/2\mathbb{Z}, + \rangle$.

(ii) Prove that χ is a homomorphism of $\langle S, \circ \rangle$ (even/odd function under composition) to $\langle \mathbb{Z}/2\mathbb{Z}, \cdot \rangle$.

2.23 Let $f : \mathbb{R} \to S^1$ where $f(\theta) = e^{i\theta} = \cos\theta + i \sin\theta$. Decide if f is a homomorphism from the addition of real numbers to the multiplication of complex numbers.

2.24 Show that the determinant is a homomophism $\det : M_n(\mathbb{R}) \to \mathbb{R}$, for $n = 2, 3$. (Later we shall explain how to do this for all n without resorting to computations.)

We are now able to state and prove the main theorem in all of algebra. It may appear obvious given our leadin, but its value should not be underestimated. We will call it *the Fundamental Homomorphism Theorem* to place it in a league with other power theorems such as the Fundamental Theorem of Calculus and the Fundamental Theorem of Algebra (which incidently has nothing to do with Algebra but rather everything to do with analysis and so it is not the subject of this course). This theorem can be stated in other ways, such as we do in the corollary to follow.

Theorem 2.2.6 (The Fundamental Homomorphism Theorem)**.** *Let S and T be sets with n-ary operations $[s_1, \dots, s_n]$ and $[t_1, \dots, t_n]$ respectively. If $f : S \to T$ is a homomorphism (with respect to the given operations) then the induced partition S/f is a quotient of S and the induced functions $\check{f} : S \to S/f$ and $\hat{f} : S/f \to T$ are also homomorphisms.*

Proof. We have already seen that every function $f : S \to T$ determines a surjective function $f : S \to \operatorname{im} f$ and that surjective functions detemine partitions accoding to their fibers, i.e $S/f = \{f^{-1}(t) : t \in \operatorname{im} f\}$. So now we must verify that S/f is a quotient. It will suffice to show that the associated equivalence relation, namely that $s \equiv s'$ if, and only if, $f(s) = f(s')$, is a congruence for $[s_1, \dots, s_n]$. So let us suppose that $s_1, \dots, s_n$ and $s'_1, \dots, s'_n \in S$ such that for each i, $s_i \equiv s'_i$. That is, $f(s_i) = f(s'_i)$. We must show that $[s_1, \dots, s_n] \equiv [s'_1, \dots, s'_n]$. This is seen because f is a homomorphism and therefore

$$f([s_1, \dots, s_n]) = [f(s_1), \dots, f(s_n)] = [f(s'_1), \dots, f(s'_n)] = f([s'_1, \dots, s'_n]).$$

Thus, we have verified that $\equiv$ is a congruence for $[x_1, \dots, x_n]$. By Proposition-2.2.4, S/f is a quotient of $\langle S, [\dots] \rangle$.

The rest follows by considering the defintions of $\hat{f}$ and $\check{f}$. □

Definition 2.2.7. An *isomorphism* is an invertible homomorphism whose inverse is a homomorphism.

Lemma 2.2.8. *The inverse of an invertible homomorphism is a homomorphism.*

Proof. Let $f : S \to T$ be a homomorphism (with respect to operations $[s_1, \dots, s_n]$ on S and $[t_1, \dots, t_n]$ on T). Now let $t_1, \dots, t_n \in T$. As f is invertible there are unique $s_1, \dots, s_n \in S$ such that $f(s_i) = t_i$. Therefore, as $f(f^{-1}(t)) = t$, for any $t \in T$, it follows that

$$\begin{aligned} f(f^{-1}([t_1, \dots, t_n])) &= [t_1, \dots, t_n] \\ &= [f(s_1), \dots, f(s_n)] \\ &= f([s_1, \dots, s_n]) \\ &= f([f^{-1}(t_1), \dots, f^{-1}(t_n)]). \end{aligned}$$

Observe the third line used the assumption that f is a homomorphism. So we have shown that

$$\begin{aligned} f(f^{-1}([t_1, \dots, t_n])) &= f([f^{-1}(t_1), \dots, f^{-1}(t_n)]); \\ f^{-1}(f(f^{-1}([t_1, \dots, t_n]))) &= f^{-1}(f([f^{-1}(t_1), \dots, f^{-1}(t_n)])); \\ f^{-1}([t_1, \dots, t_n]) &= [f^{-1}(t_1), \dots, f^{-1}(t_n)]. \end{aligned}$$

In particular, f^{-1} is a homomorphism. □

Proposition 2.2.9. *Isomorphism is an equivalence relation for the class of all sets with n-ary operators.*

2.25 Isomorphism an Equivalence Prove Proposition-2.2.9.

It may seem that every function that is invertible and has some useful property will then produce the same property for the inverse. Hence, Lemma-2.2.8 seems unsurprising. However, this is usually not the case, for example, consider continuity.

Example 2.2.10. Let $S^1 = \{(\cos\theta, \sin\theta) : \theta \in [0, 2\pi)\}$ be usual unit circle. Let $f : [0, 2\pi) \to S^1$ be the function $f(\theta) = (\cos\theta, \sin\theta)$. It follows that f is invertible and continuous. However, f^{-1} is *not* continuous.

Proof. The only important point to understand is that f^{-1} is not continuous at $(1, 0)$. That is, as we approach $(1, 0)$ from above we have $\theta \to 0$. But when we approach $(1, 0)$ from below, $\theta \to 2\pi$. Hence the right-hand limit of f^{-1} near $(1, 0)$ is 0 but the left-hand limit is 2π and so there is no overall limit at $(1, 0)$. Therefore, f^{-1} cannot be continuous at $(1, 0)$. □

Corollary 2.2.11 (The First Isomorphism Theorem)**.** *Under the hypothesis of Theorem-2.2.6, if f is surjective then $\hat{f}$ is an isomorphism.*

Proof. Use the Fundamental Homomorphism Theorem with Lemma-2.2.8. □

2.3 Kernels

We finally encounter on one of the main attractions to groups over all other σ-algebras.

Before considering the following result recall that for a function $f : A \to B$, the fibers of f are denoted $f^{-1}(b) = \{a \in A : f(a) = b\}$ and f induces a partion $A/f = \{f^{-1}(f(a)) : a \in A\}$. The Fundamental Homomorphism Theorem demonstrated that if $f : G \to H$ is a homomorphism for sets G and H with n-ary operators, then the partition G/f is also a quotient and the function $\hat{f} : G \to G/f$ given by $\hat{f}(g) = f^{-1}(f(g))$ is a homormophism. We now consider the effect of considering groups rather than arbitrary sets with operators. The result is stricking and pleasing.

Theorem 2.3.1. *If G is a group and $f : G \to H$ is a homomorphism, then*

(i) The fiber $f^{-1}(f(1))$ is a group using the same operations as G, i.e. the same product, inverses, and identity.

(ii) $G/f = \{gf^{-1}(f(1)) : g \in G\}$ where $gf^{-1}(f(1)) = \{gk : k \in f^{-1}(f(1))\}$.

Proof. Set $K = f^{-1}(f(1)) = \{g \in G : f(g) = f(1)\}$.

(i). As G is a group it comes with three operations: the binary operation $\cdot : G^2 \to G$, the unary operation $^{-1} : G \to G$, and the nullary (constant) operation $1 : G^0 \to G$. As K is a subset of G we can look at the restriction of each of these three functions to see if they produce operations on K. First, if $x, y \in K$ then $f(x) = f(1) = f(y)$. Therefore, $f(x \cdot y) = f(x) \cdot f(y) = f(1) \cdot f(1) = f(1 \cdot 1) = f(1)$. Thus, $x \cdot y \in K$. That is, the function $\cdot$ now restricts to $\cdot : K^2 \to K$ and so K has a binary operation. Secondly, $f(1) = f(1)$ so $1 \in K$. Thirdly, for every $x \in K$, $f(x) = f(1)$ and so $f(x)^{-1} = f(1)^{-1}$ (which makes sense because we now the image of f is a group and so inverses exist for every image element) and so $f(x^{-1}) = f(1^{-1}) = f(1)$. Thus, $x^{-1} \in K$ and so $^{-1} : K \to K$. Therefore K is a group.

(ii). Fix a fiber $f^{-1}(f(g)) \in G/f$. We will show $f^{-1}(f(g)) = gf^{-1}(f(1))$. First, for every $x \in f^{-1}(f(g))$, $f(x) = f(g)$. Therefore,

$$f(g^{-1}x) = f(g)^{-1}f(x) = f(g)^{-1}f(g) = f(g^{-1}g) = f(1).$$

Thus, $g^{-1}x \in f^{-1}(f(1))$. Furthermore, $x = (gg^{-1}x) = g(g^{-1}x) \in gf^{-1}(f(1))$. Therefore $f^{-1}(f(g)) \subseteq gf^{-1}(f(1))$. Second, for all $y \in gf^{-1}(f(1))$, $y = gm$ where $m \in f^{-1}(f(1))$, that is, $f(m) = f(1)$. Thus, $f(y) = f(gm) = f(g)f(m) = f(g)f(1) = f(g \cdot 1) = f(g)$. Thus $y \in f^{-1}(f(g))$ and so $gf^{-1}(f(1)) \subseteq f^{-1}(f(g))$.

So we see

$$G/f = \{f^{-1}(f(g)) : g \in G\} = \{gf^{-1}(f(1)) : g \in G\}.$$

□

Definition 2.3.2. If G is a group and $f : G \to H$ is a homomorphism then the *kernel* of f is $f^{-1}(f(1))$ and denoted by $\ker f$. We also write $G/\ker f$ for G/f. The elements of $G/\ker f$ are called the *cosets of* $\ker f$.

2.26 Prove under the condition of Theorem-2.3.1, we also have $G/f = \{f^{-1}(f(1))g : g \in G\}$ where $f^{-1}(f(1))g = \{kg : k \in f^{-1}(f(1))\}$.

2.27 Find the kernel of the homomorphism $f : \mathbb{Q} \to S^1$, $f(a/b) = e^{2\pi ia/b}$.

2.28 Find the kernel of the homomorphism $f : \mathbb{Z} \to S^1$, $f(n) = e^{2\pi in/12}$.

2.29 Let $\mathrm{GL}_2(\mathbb{Q})$ be the set of invertible (2×2)-matrices with entries in $\mathbb{Q}$. Find the kernel of the homomorphism $f : \mathrm{GL}_2(\mathbb{Q}) \to \mathbb{Q}^\times$ (i.e. the rationals without zero and with multiplication as the operation) where $f(A) = \det A$.

2.4 Further exercises

2.30 Decide if the subsets relation $\subseteq$ is an equivalence relation on the class of all sets.

2.31 Trivial operator. Explain what is meant when we say "Suppose $\langle S, \cdot\rangle$ satisfies the identity axiom and ....blah blah blah.... Thus, $S = \{1\}$."

2.32 Say that two differentiable real functions f and g are *parallel*, denoted $f \| g$, if $\frac{df}{dx} = \frac{dg}{dx}$.

(i) Is parallel is an equivalence relation?

(ii) Is parallel a congruence for addition of differentiable real functions?

(iii) Is parallel a congruence for multiplication of differentialble real functions?

2.33 Let $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ be the xy-plane with vector addition. Prove that for any (2×1)-matrix $\begin{bmatrix} a \\ b \end{bmatrix}$, the function $f(x, y) = [x, y] \begin{bmatrix} a \\ b \end{bmatrix}$ is a homomorphism from $\langle \mathbb{R}^2, +\rangle$ to $\langle \mathbb{R}, +\rangle$.

2.34 Unique trivial operator. Prove that if $\langle S, \cdot\rangle$ and $\langle T, \cdot\rangle$ are satisfy the identity axiom and both sets S and T have size 1 then there is an isomorphism from S to T.

2.35 Unital homomorphisms. True or False? Suppose that $f : \langle S, *\rangle \to \langle T, \#\rangle$ is a homomorphism and that both S and T have identities. Explain what it means when we say $f(1) = 1$. Is it true that $f(1) = 1$?

2.36 Is the transpose function on square matrices a homomorphism of the addition of matrices? What about the multiplication of matrices?

2.37 Fix three sets $\langle S, \{s_1, \ldots, s_n\}\rangle$, $\langle T, [t_1, \ldots, t_n]\rangle$, and $\langle U, /u_1, \ldots, u_n/\rangle$ with n-ary operations. Prove that if $f : S \to T$ is a homomorphism and $g : T \to U$ is a homomorphism, then $g \circ f$ is a homomorphism.

2.38 Prove that if $\mathscr{Q}$ is a quotient of $\langle S, \{s_1, \ldots, s_n\}\rangle$ and $f : S \to T$ is a surjective homomorphism to $\langle T, [t_1, \ldots, t_n]\rangle$, then $f(\mathbb{Q}) = \{\{f(s) : s \in Q\} : Q \in \mathbb{Q}\}$ is a quotient of $\langle T, [t_1, \ldots, t_n]\rangle$.

2.39 Polarizitaion Let $\langle \mathbb{R}, + \rangle$ be the real numbers under ordinary addition of decimal numbers. If $\{s\}$ is a unary operator on $\mathbb{R}$ then we can produce a new binary operation $\star$ by setting

$$a \star b = \{a + b\} - \{a\} - \{b\} \qquad (\forall a, b \in \mathbb{R}).$$

Prove that if $\{s\} = s^2$, then $\langle \mathbb{R}, \star \rangle$ is isomorphic to $\langle \mathbb{R}, \cdot \rangle$.

2.40 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = (x, y) \begin{bmatrix} -1 \\ 3 \end{bmatrix}$.

(i) Prove that f is surjective.

(ii) Describe in words what the equivalence relation on $\mathbb{R}^2$ is produced by f.

(iii) Decide if the the equivalence relation is a congruence of $\langle \mathbb{R}^2, + \rangle$ (the usual addition of vectors).

2.41 Let $F = \mathbb{Q} \times \mathbb{Q}$. We know that the usual coordinate comparision $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$ makes an equivalence relation on F. If we define a binary operation as follows

$$(a, b) \cdot (c, d) = (ac + 2bd, ad + bc);$$

then show that $=$ is a congruence for this operation.

Chapter 3

Direct Products

Motivation

In this chapter we study the second fundamental method to create new algebras in a variety, the method of direct products. Direct products have undergone several reinterpretations. The first recorded us of the name *direct product* was in 1886 when Otto Hölder's classification of various small groups. Later in 1909 Wedderburn published the first proof that direct products are essentially unique. That encouraged mathematicians to consider them as a useful means to exam groups. Remak, Krull, Schmidt, Fitting, Kurosh, Azumaya, and Ore each offered stronger and simpler proofs of the results. Levi and Birhkoff saw the value of direct products to varities and finally Ellenberg and MacLane fit the entire construction into their much more general system of categories.

3.1 Quotients and Free σ-algebras

Further Reading: Jacobson §1.11.

At last we create something universal (which is the point of Universal Algebra)! The next theorem might appear easy to prove but that is a result of good forethought in choosing definitions.

Theorem 3.1.1. *Let G and H be sets with operators of signature σ. Let Φ be a set of σ-equations. If $G \in \mathfrak{V}(\Phi)$ and $f : G \to H$ is an epimorphism (for all the operations in σ) then $H \in \mathfrak{V}(\Phi)$.*

Proof. For each $\phi(x_1, \ldots, x_\ell) = \gamma(x_1, \ldots, x_\ell)$ be a sentence in Φ. For all $h_1, \ldots, h_\ell \in H$, as f is surjective there are $g_1, \ldots, g_\ell \in G$ such that for each i, $f(g_i) = h_i$. First, because f is a homomorphism for every operation in σ, and ϕ is a conjunction of operations in σ, it follows that

$$f(\phi(g_1, \ldots, g_\ell)) = \phi(f(g_1), \ldots, f(g_\ell)).$$

A same holds if we replace ϕ with γ. Hence,

$$\begin{aligned} \phi(h_1, \ldots, h_\ell) &= \phi(f(g_1), \ldots, f(g_\ell)) \\ &= f(\phi(g_1, \ldots, g_\ell)) \\ &= f(\gamma(g_1, \ldots, g_\ell)) \\ &= \gamma(f(g_1), \ldots, f(g_\ell)) \\ &= \gamma(h_1, \ldots, h_\ell). \end{aligned}$$

Notice in the middle we exchanged ϕ for γ which follows as we know $G \in \mathfrak{V}(\Phi)$. Thus, $H \in \mathfrak{V}(\Phi)$. □

We often communicate Theorem-3.1.1 by saying that "Varieties are closed to homomorphic images." The word *closed* simply means that if an object has the property we mention then the set we mention contains that object. In another context we might say "The integers are closed to subtract but the natural numbers are not." As a community we have resisted the temptation to use the word "open" to describe sets and classes that are not closed to a specified property. Thus, we have avoided suggesting there is a topological process involved.

3.1 Show that a variety is closed to quotients.

3.2 Let $\sigma = \{\cdot, 1\}$ and $\Phi = \{x \cdot 1 = x = 1 \cdot x\}$ be the usual equation defining the identity axiom. Without using Theorem-3.1.1, show directly that if G and H are sets with binary operations, $G \in \mathfrak{V}(\Phi)$, and $f : G \to H$ is an epimorphism, then $H \in \mathfrak{V}(\Phi)$.

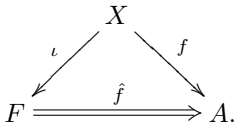
3.3 Let $\sigma = \{\cdot, 1\}$ and $\Phi = \{x \cdot y = y \cdot x\}$ be the usual equation for commutativity. Without using Theorem-3.1.1, show directly that if G and H are sets with binary operations, $G \in \mathfrak{V}(\Phi)$, and $f : G \to H$ is an epimorphism, then $H \in \mathfrak{V}(\Phi)$.

We now understand that quotients and homomorphic images of an element in a variety are guaranteed to lie back in the variety. So thinking universally we ask, is there some member F in a variety $\mathfrak{V}$ that is so large that all the members in $\mathfrak{V}$ are simply quotients of it? If so, we could get by studying just this one master member. The problem however is in Russell's Paradox. There is not set of all sets and so we cannot generally expect that we an construction a set F large enough to have surjections onto all other sets in $\mathfrak{V}$. But the idea is not entirely invalid. Instead we ask, can we make an F which covers all members of $\mathfrak{V}$ up to a fixed size?

Definition 3.1.2. An algebra F with a function $\iota : X \to F$ from a set X is called *free on* X, with respect to a variety $\mathfrak{V}$, if whenever $A \in \mathfrak{V}$ and $f : X \to A$ is a function, there is a unique homomorphism $\hat{f} : F \to A$ such that

$$\forall x \in X, \qquad f(x) = \hat{f}(\iota(x)).$$

It often helps to visualize the functions involved.



We use a double-bared function to indicate that this function is unique with respect to the others and is implied to exist by the others. The idea that a combination of functions is enough to predict the existence and uniqueness of another is known as a *Universal Mapping Property*. Notice that X is only a set, not an algebra, so it makes no sense to ask if ι and f are homomorphisms. Indeed they are only functions. However, $\hat{f}$ is homomorphism.

For example. Suppose we look at the variety $\mathfrak{V} = \mathfrak{V}(x + y = y + x,\ x + (y + z) = (x + y) + z)$, i.e. the variety of signature $\{+\}$ where $+$ satisfies the associative and commutative laws. Notice $\langle \mathbb{N}, + \rangle$ is in this variety, as are $\langle \mathbb{Z}/N\mathbb{Z}, + \rangle$ for every $N \in \mathbb{Z}$, $\langle \mathbb{Q}, + \rangle$, $\langle \mathbb{R}, \cdot \rangle$, and many other objects. Let us take $X = \{a\}$ and $\iota : X \to \mathbb{N}$ to be $f(a) = 1$. We will demonstrate the $\mathbb{N}$, together with ι, is free on X (in the variety $\mathfrak{V}$).

First consider an example. Suppose that$f : X \to \mathbb{Z}/12\mathbb{Z}$ be the function $f(a) = 5$. We must decide if there is a homomorphism $\hat{f} : \mathbb{N} \to \mathbb{Z}/12\mathbb{Z}$ which behaves like $f(x) = \hat{f}(\iota(x))$ for each $x \in X$. As $X = \{a\}$ we have very little to test. We simply need

$$5 = f(a) = \hat{f}(\iota(a)) = \hat{f}(1).$$

It may seem impossible that by specifying only one output for $\hat{f}$ we now hope to describe where each of the infinitely many inputs in $\mathbb{N}$ are headed. However, we must recall that $\hat{f}$ is to be a homomorphism! Thus, we are forced to make other outputs agree with our first choice that $\hat{f}(1) = 2$. So we find:

$$\begin{aligned}
\hat{f}(2) &\equiv \hat{f}(1+1) \equiv \hat{f}(1) + \hat{f}(1) \equiv 5 + 5 \equiv 10 \pmod{12};\\
\hat{f}(3) &\equiv \hat{f}(2+1) \equiv \hat{f}(2) + \hat{f}(1) \equiv 10 + 5 \equiv 3;\\
\hat{f}(n+1) &\equiv \hat{f}(n+1) \equiv \hat{f}(n) + \hat{f}(1) \equiv 2n + 2 \equiv 2n + 2 \equiv 2(n+1).
\end{aligned}$$

So in the end we had not choice in where the remaining values of $\mathbb{N}$ are sent under $\hat{f}$. So we have defined a function $\hat{f} : \mathbb{N} \to \mathbb{Z}/12\mathbb{Z}$ where $\hat{f}(n) = 2n$. This in fact a homomorphism because

$$\hat{f}(n+m) \equiv 2(n+m) \equiv 2n + 2m \equiv \hat{f}(n) + \hat{f}(m) \pmod{12}.$$

Now we consider the general case. Suppose that $\langle A, + \rangle \in \mathfrak{V}$ and $f : \{a\} \to A$ is a function. Define $\hat{f} : \mathbb{N} \to A$ as follows:

$$\begin{aligned} \hat{f}(1) &= f(a) \\ \hat{f}(n+1) &= \hat{f}(n) + f(a) \qquad\qquad (\forall n \in \mathbb{N}). \end{aligned}$$

This does define a function. Furthermore, it defines a homomorphism, however, the proof of that is a somewhat involved induction.

To introduce free objects in semigroups, monoids, and groups we have to first describe what algebraist refer to as words. Despite the following tedious definition, as the example to follow shows, the idea is quite natural.

Definition 3.1.3. Given a set $\mathtt{X}$, a *word* of length $n \in \mathbb{N}$ is a function $w : \{1, \dots, n\} \to \mathtt{X}$. We call $\mathtt{X}$ the alphabet of the word w. The *concatenation* of words $w : \{1, \dots, n\} \to \mathtt{X}$ and $u : \{1, \dots, m\} \to \mathtt{X}$ is the word $wu : \{1, \dots, n+m\} \to \mathtt{X}$ defined by:

$$wu(i) = \begin{cases} w(i) & i \leq n; \\ u(i-n) & i > n. \end{cases}$$

The empty-word is the function $\epsilon : \varnothing \to \mathtt{X}$ (it has no points in the domain).

Example 3.1.4. For the set $\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \dots, \mathtt{x}, \mathtt{y}, \mathtt{z}$ a word can be represented as a common word. For instance, `alleycat` is a word of length 8. The implied function is $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \mathtt{a} & \mathtt{l} & \mathtt{l} & \mathtt{e} & \mathtt{y} & \mathtt{c} & \mathtt{a} & \mathtt{t} \end{pmatrix}$. Notice this is the concatenation of the words `alley` and `cat`. The concatenation of `cat` and `alley` is `catalley` which is obviously not the same.

Unlike language, words in algebra do not need to be in the dictionary they can be any assortment of letters, e.g. `adkfj` is a word of length 5. Notice, `Häagen Daz` is not a word over the set $\mathtt{a}, \mathtt{b}, \mathtt{c}, \dots, \mathtt{x}, \mathtt{y}, \mathtt{z}$ because `H`, `ä`, `D` and the space are not in the set. This could become a word over a different set.

We usually write words $w : \{1, \dots, n\} \to \mathtt{X}$ simply as $w(\mathtt{X})$. For example, $w(a, b, c, \dots, x, y, z) = alleycat$ or in general if $\mathtt{X} = \{x_1, x_2, x_3\}$ we might describe a word as $w(x_1, x_2, x_3) = x_1x_2x_1x_3x_2$. This encourages us to think of the letters in the alphabet $\mathtt{X}$ as variables which might later substitute with values.

Theorem 3.1.5. *In the variety $\mathfrak{S}$ of semigroups, for each set $\mathtt{X} = \{x_1, x_2, \dots\}$, the set $Fr_{\mathfrak{S}}[\mathtt{X}]$ of non-empty words in $\mathtt{X}$ equipped with concatenation as a product is a free semigroup on $\mathtt{X}$. Furthermore, if S is a semigroup and $f : \mathtt{X} \to S$ is a function, i.e. for each $x_i \in \mathtt{X}$ we assign $f(x_i) = s_i \in S$, then define the required unique homomorphism $\hat{f} : Fr_{\mathfrak{S}}[\mathtt{X}] \to S$ as follows:*

$$\hat{f}(w(x_1, x_2, \dots) = w(f(x_1), f(x_2), \dots) = w(s_1, s_2, \dots)$$

for every word $w(x_1, x_2, \dots) \in Fr_{\mathfrak{S}}[\mathtt{X}]$.

Proof. First we observe that $F := Fr_{\mathfrak{S}}[\mathtt{X}]$ is a semigroup because given non-empty words $w : \{1, \dots, \ell\} \to \mathtt{X}$, $u : \{1, \dots, m\} \to \mathtt{X}$, and $v : \{1, \dots, n\} \to \mathtt{X}$, we

have

$$\begin{aligned}((wu)v)(s) &= \begin{cases} (wu)(s) & 1 \leq s \leq \ell + m \\ v(s-(\ell+m)) & s > \ell + m \end{cases} \\ &= \begin{cases} w(s) & 1 \leq s \leq \ell \\ u(s-\ell) & \ell < s \leq \ell + m \\ v(s-(\ell+m)) & s > \ell + m \end{cases} \\ &= \begin{cases} w(s) & 1 \leq s \leq \ell \\ (uv)(s-\ell) & \ell < s \end{cases} \\ &= (w(uv))(s).\end{aligned}$$

Although necessary, the proof of associativity is also evident by considering examples, say if $w(\mathtt{a},\mathtt{b},\mathtt{c},\ldots,\mathtt{x},\mathtt{y},\mathtt{z}) = \mathtt{my}$, $u(\mathtt{a},\mathtt{b},\mathtt{c},\ldots,\mathtt{x},\mathtt{y},\mathtt{z}) = \mathtt{alley}$, and $v(\mathtt{a},\mathtt{b},\mathtt{c},\ldots,\mathtt{x},\mathtt{y},\mathtt{z}) = \mathtt{cat}$, then

$$(wu)v = (\mathtt{myalley})\mathtt{cat} = \mathtt{myalleycat} = \mathtt{my}(\mathtt{alleycat}) = w(uv).$$

Next we demonstrate that F is free on $\mathtt{X}$, which means we confirm F satisfies the Universal Mapping Property for Free Algebras (UMP4FA). Let S be a semigroup and $f : \mathtt{X} \to S$ a function. Suppose there is a homomorphism $\hat{f} : F \to S$ such that for all $x \in \mathtt{X}$, $\hat{f}(x) = f(x)$ (recall $x \in F$ as x is a word in the alphabet $\mathtt{X}$). If $w : \{1,\ldots,n\} \to \mathtt{X}$ is a word of length n in F then $w = w(1)w(2)\cdots w(n)$ – i.e. the word is the concatenation of its letters, for example $\mathtt{cat} = \mathtt{c}\cdot\mathtt{a}\cdot\mathtt{t}$. Each $w(i) \in \mathtt{X}$ and so $\hat{f}(w(i)) = f(w(i))$. Together we find

$$\hat{f}(w) = \hat{f}(w(1)w(2)\cdots w(n))\hat{f}(w(1))\hat{f}(w(2))\cdots\hat{f}(w(n))f(w(1))f(w(2))\cdots f(w(n)).$$

Said differently, if there is a homomorphism $\hat{f} : F \to S$ such that for all $x \in \mathtt{X}$, $\hat{f}(x) = f(x)$, then for every word $w(x_1, x_2, \ldots)$

$$\hat{f}(w(x_1, x_2, \ldots) = w(f(x_1), f(x_2), \ldots).$$

Now that we have confirmed that at most one homomorphism can exist with the property we seek, it remains to claim that $\hat{f}$ actually exists. Clearly $\hat{f}$ is a function so it remains to show $\hat{f}$ is a homomorphism.

$$\begin{aligned}\hat{f}(w(x_1,x_2,\ldots)u(x_1,x_2,\ldots)) &= \hat{f}((wu)(x_1,x_2,\ldots)) \\ &= (wu)(f(x_1),f(x_2),\ldots) \\ &= w(f(x_1),f(x_2),\ldots)u(f(x_1),f(x_2),\ldots) \\ &= \hat{f}(w(x_1,x_2,\ldots))\hat{f}(u(x_1,x_2,\ldots)).\end{aligned}$$

Hence $\hat{f}$ is a homomorphism and our proof is complete. □

Corollary 3.1.6. *In the variety $\mathfrak{M}$ of monoids, for each set $\mathtt{X}$, the set $Fr_{\mathfrak{M}}[\mathtt{X}] = Fr_{\mathfrak{S}}[\mathtt{X}] \cup \{\epsilon\}$ of words in $\mathtt{X}$, a form 'empty-word' ϵ, with concatenation as a product (where also $w\epsilon = w = \epsilon w$ for all $w \in Fr_{\mathfrak{S}}[\mathtt{X}]$) is a free monoid on $\mathtt{X}$.*

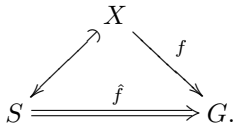
Proof. Use the unique homomorphism guaranteed by the Universal Mapping Property of Free Semigroups and then send the empty-word in $Fr_{\mathfrak{M}}[\mathtt{X}]$ to the identity in the target monoid. □

Definition 3.1.7. For a word $w : \{1,\ldots,n\} \to \mathtt{X}$, let $w^{rev} : \{1,\ldots,n\} \to \mathtt{X}$ be $w^{rev}(i) = w(n-i)$.

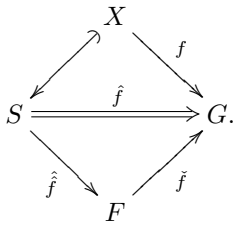
The reversal of a word is obvious with examples. If $w(\mathtt{a},\mathtt{b},\mathtt{c},\ldots,\mathtt{x},\mathtt{y},\mathtt{z}) = \mathtt{alley}$ then $w^{rev}(\mathtt{a},\mathtt{b},\mathtt{c},\ldots,\mathtt{x},\mathtt{y},\mathtt{z}) = yella$.

Corollary 3.1.8. *In the variety $\mathfrak{G}$ of groups, for each set $\mathtt{X}$, the set $Fr_{\mathfrak{G}}[\mathtt{X}] = Fr_{\mathfrak{M}}\mathtt{X} \cup \mathtt{X}^{-1}]$ of words in $\mathtt{X}$ an in the formal copy $\mathtt{X}^{-1}$ of X, modulo the congruence relation $w(\mathtt{X})w(\mathtt{X})^{rev} \equiv \epsilon$, is a free group on $\mathtt{X}$.*

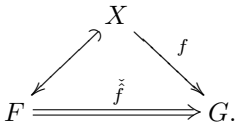
Proof. Let G be a group and $f : \mathtt{X} \to G$ a function. From the Universal Mapping Property for Free Semigroup $S = Fr_{\mathfrak{S}}[\mathtt{X} \cup \mathtt{X}^{-1}]$ we have the diagram:



As $\equiv$ is a congruence relation on the semigroup S so there is an induced quotient $F = Fr_{\mathfrak{G}}[\mathtt{X}]$ and also by the Fundamental Homomorphism Theorem there are homomorphisms $\hat{\hat{f}}$ and $\check{\hat{f}}$ as in the diagram below.



The congruence relation we have imposed makes the quotient have inverses, so F is a group. In particular we obtain



And so F is a free group. □

Definition 3.1.9. Fix a variety $\mathfrak{V}$. For a $G \in \mathfrak{V}$, G is said to be *generated by a subset* X there is an $F \in \mathfrak{V}$ which is free on X and such that the homomorphism $\hat{f} : F \to G$ given by the inclusion $f : X \to G$ (i.e. $f(x) = x$) is an epimorphism.

3.2 Direct Products

We have used operations on sets and now we begin to involve operations on varieties. The most natural operation is that of a direct product. It hinges on the concept of a Cartesian product of sets. The best known example is the xy-plane which was introduced by Rene des Cartes as a scheme to render functions $f : \mathbb{R} \to \mathbb{R}$ as images by plotting the points $\{(x, f(x) : x \in \operatorname{dom} f\}$. The name Cartesian product derives from his invention. We start there.

3.2.1 Cartesian Products

For sets A and B we write $A \times B$ for the set of ordered pairs. So the xy-plane is denoted $\mathbb{R} \times \mathbb{R}$ or sometimes $\mathbb{R}^{\times 2}$ or simply $\mathbb{R}^2$. Later with a set C we can ask for $(A \times B) \times C$ or $A \times (B \times C)$. These sets are roughly the same idea but are nevertheless not equal. For example, the first has elements of the form $((a, b), c)$ whereas the second set has elements of the form $(a, (b, c))$. There is an invertible function $(A \times B) \times C \to A \times (B \times C)$, namely $((a, b), c) \mapsto (a, (b, c))$; hence,

in that way the Sets are strongly related. This leads many authors to declare somewhat informally that the Cartesian product of sets is associative. This is not entirely true becuase we just mentioned $A \times (B \times C) \neq (A \times B) \times C$, it is only true if we involve the equivalence relation $\sim$ on sets which says, $X \sim Y$ if and only if there is an invertible function $f : X \to Y$. In that notation we can legitimately write

$$A \times (B \times C) \sim (A \times B) \times C.$$

However, $\sim$ goes too far in most cases. For example, there is an invertible function from $\mathbb{N}$ to $\mathbb{Z}$ (e.g. $f(n) = n/2$ if n is even and $f(n) = (1-n)/2$ if n is odd.) Indeed that allows us to make an invertible function $\mathbb{N} \times \mathbb{N} = \mathbb{Z} \times \mathbb{Z}$. Likewise $\mathbb{R}$ is in one-to-one correspondence with $(0,1)$ but yet again we do not seriously consider the entire xy-plane $\mathbb{R}^{\times 2}$ as the same as the unit square $(0,1)^{\times 2}$. So we need a more subtle solution.

3.4 Construct an invertible function $f : \mathbb{R} \to (0,1)$ and also $g : \mathbb{R}^2 \to (0,1)^2$.

The way out of the associativity problem for Cartesian products is to introduce n-ary Cartesian products on sets.

$$\begin{aligned} A \times B &= \{(a,b) : a \in A, b \in B\} \\ A \times B \times C &= \{(a,b,c) : a \in A, b \in B, c \in C\} \\ A \times B \times C \times D &= \{(a,b,c,d) : a \in A, b \in B, c \in C, d \in D\} \\ &\vdots \end{aligned}$$

There are technical difficulties in doing this. First, what is meant by (a,b,c,d) and second, what if we want a product of an infinite number of sets? Using functions this becomes evident.

Definition 3.2.1. Let $\mathcal{S}$ be the class of sets.

(i) A *family* $\mathcal{F}$ (or *indexed-set*) is a set I and a function $F : I \to \mathcal{S}$. We usually write F_i (instead of the usual $F(i)$) and write $\mathcal{F} = \{F_i : i \in I\}$.

(ii) To families $\mathcal{F} = \{F_i : i \in I\}$ and $\mathcal{G} = \{G_j : j \in J\}$ are said to be equal when there is an invertible function $t : I \to J$ such that for all $i \in I$, $F_i = G_{t(i)}$.

We introduced families in a somewhat pompous manner but the concept is basic. With sets we have $\{A, A, B, A\} = \{A, B\}$ because equality of sets occurs whenever two sets have the same members. However families, even though they are denoted as sets, are actually functions. Therefore the family $\mathcal{F} = \{A = F_1, A = F_2, B = F_3, A = F_4\}$ is *not* equal to the family $\mathcal{G} = \{A = G_1, B = G_2\}$ because the functions implied here are rather distinct. In the first we have $F : \{1,2,3,4\} \to \{A, B\}$ and in the second we have $G : \{1,2\} \to \{A, B\}$. So using families we have the option to repeat terms as often as we like. The equivalence of families shows that we do not care about the order of the terms, just that the count be the same. This still introduces certain oddities when the index sets are infinite but we leave that for future discussion.

Remark 3.2.2. Recall that it is sufficient in most setting to give the name of a function and not to describe its domain or codomain. This brevity is especially welcome (though uncommon) with families. For example, it is permissible to write:

> Given a family $\mathcal{F} = \{F_i : i \in I\}$, for each $i \in I$, F_i...

On the other-hand, there is a modest elegance in writing

Given a family $\mathcal{F}$, for each $F \in \mathcal{F}$,...

However, it is important to stress in the latter notation that $F \in \mathcal{F}$ means to run through all the values of the function including any repeated values. This simply avoids extra subscripts which soon become a nuisances. The correct understanding that $F \in \mathcal{F}$ means to include multiple copies is made clear by having first indicated that $\mathcal{F}$ is a *family* and not simply a set.

Definition 3.2.3. The *Cartesian product* $\prod \mathcal{F} = \prod_{F\in\mathcal{F}} F$ of a family $\mathcal{F}$ is the set of all function $f : \mathcal{F} \to \bigcup_{F\in\mathcal{F}} F$ such that for every $F \in \mathcal{F}$, $f(F) \in F$. For each $F \in \mathcal{F}$, the *projection function* $\pi_F : \prod \mathcal{F} \to F$ is defined by $\pi_F(f) = f(F)$.

When a family $\mathcal{F}$ has a small index set, e.g. if $I = \{1, 2, 3\}$, we usually denote $\prod_{i\in I} F_i$ by $F_1 \times F_2 \times F_3$. In the usual coordinate notation the three projections here would be $\pi_1(a, b, c) = a$, $\pi_2(a, b, c) = b$, and $\pi_3(a, b, c) = c$.

These definitions can be tedious and so we take a moment to explore it in a very gentle setting.

Example 3.2.4. Fix sets $A = \{1, 2, 3\}$ and $B = \{\alpha, \beta\}$, and $C = \{\star, \circ\}$. Let $\mathcal{F} = \{A, B, C, B\}$ be a family (the implied index set is $I = \{1, 2, 3, 4\}$). Then $\prod \mathcal{F}$ is the set of functions

$$f : \{1, 2, 3, 4\} \to (A \cup B \cup C \cup B) = \{1, 2, 3, \alpha, \beta, \star, \circ\} \tag{3.1}$$

where $f(1) \in \{1, 2, 3\}$, $f(2) \in \{\alpha, \beta\}$, $f(3) \in \{\star, \circ\}$, and $f(4) \in \{\alpha, \beta\}$. For instance:

(i) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & \beta & \star & \alpha \end{pmatrix}$ is a function in $\prod \mathcal{F}$. This function represents the 4-tuple $(2, \beta, \star, \alpha) \in A \times B \times C \times B$.

(ii) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & \circ & \star & \alpha \end{pmatrix}$ is a function $\{1, 2, 3, 4\} \to \{1, 2, 3, \alpha, \beta, \star, \circ\}$ but not in $\prod \mathcal{F}$ because it assigns 2 to $\circ$; yet, $\circ \notin F_2 = \{\alpha, \beta\}$. This function represents the 4-tuple $(3, \circ, \star, \alpha)$ which is not in $A \times B \times C \times B$, rather, it lies in $A \times C \times C \times B$.

(iii) $\begin{pmatrix} 1 & 2 & 3 \\ 2 & \beta & \star \end{pmatrix}$ is a function $\{1, 2, 3\} \to A \cup B \cup C \cup B$; yet, it lies in $A \times B \times C$ and not in $A \times B \times C \times B$. It represents a 3-tuple $(2, \beta, \star)$.

(iv) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 5 & 6 & 7 \end{pmatrix}$ is not in $\prod \mathcal{F}$ as it does not describe a function $\{1, 2, 3, 4\} \to \{1, 2, 3, \alpha, \beta, \star, \circ\}$. This represents the 4-tuple $(4, 5, 6, 7)$ which is entirely unrelated to $A \times B \times C \times B$.

Example 3.2.5. (i) The usual xy-plane $\mathbb{R}^2$ is the Cartesian product of $\mathbb{R}$ with $\mathbb{R}$. More formally, the family $F : \{x, y\} \to \mathcal{S}$ with $F_x = \mathbb{R}$ and $F_y = \mathbb{R}$ has $\{f : \{x, y\} \to \mathbb{R} : f(x), f(y) \in \mathbb{R}\}$.

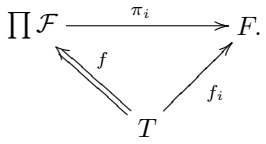
(ii) For sets X and Y, the set $\mathrm{fun}(X, Y)$ of all functions $f : X \to Y$ is also $\prod_{x\in X} Y$. The implied family here has X as the index set and for every $x \in X$ $F_x = Y$. This Cartesian product is often abreviated by Y^X. E.g. if $X = \{1, 2\}$ then $Y^X = Y \times Y$.

Theorem 3.2.6 (Universal Mapping Property for Cartesian Products)**.** *If $\mathcal{F} = \{F_i : i \in I\}$ is a family of sets, T is a set, and $\{f_i : T \to F_i : i \in I\}$ is a family of functions on sets*[1] *then there is a unique function $f : T \to \prod \mathcal{F}$ such that*

$$(\forall i \in I, \forall t \in T), \quad \pi_i(f(t)) = f_i(t).$$

[1]Recall that a function $f : A \to B$ between sets A and B is also a set, i.e. a subset of $A \times B$. So discussing families of functions between sets is permissible.

As a diagram of functions this means:



(We use the double-bar function to indicate that function is unique with respect to all others in the diagram.) The function f is often denoted $\prod_{F\in\mathcal{F}} f_i$.

Proof. We need a function $f : T \to \prod \mathcal{F}$. The elements of $\prod \mathcal{F}$ are themselves functions. Therefore such a function f takes as input an element $t \in T$ and outputs a function $f(t) : I \to \bigcup_{i\in I} F_i$. The recipe for this function is described by the equation is must satisfy. So we define:

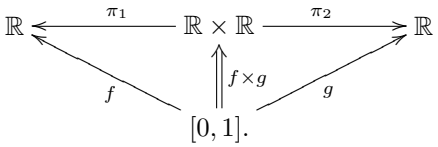
$$f(t)(i) = f_i(t) \qquad (\forall t \in T, \forall i \in I).$$

For each $t \in T$ and $i \in I$, $f_i(t) \in F_i$ (as $f_i : T \to F_i$) and so indeed $f(t) : I \to \bigcup_{i\in I} F_i$ and for each $i \in I$, $f(t)(i) \in F_i$ which shows that $f(t) \in \prod \mathcal{F}$. This explains that f is indeed a function $T \to \prod \mathcal{F}$. The property required of f is satisfied as well since

$$\pi_i(f(t)) = f(t)(i) = f_i(t) \qquad (\forall t \in T, \forall i \in I).$$

□

The Universal Mapping Property for Cartesian products (UMP4CP) may seem abstract but the concept is easy to following with a picture. Suppose we have $f(x) = x^2$ and $g(x) = x - x^3$ as functions $[0, 1] \to \mathbb{R}$. These are already enough data to demonstrate the UMP4CP. First we identify the sets involved. The functions $f, g : [0, 1] \to \mathbb{R}$ will form the family of functions we need in UMP4C. Thus, $T = [0, 1]$ and $\mathcal{F} = \{\mathbb{R}, \mathbb{R}\}$. The product we form is $\prod \mathcal{F} = \mathbb{R} \times \mathbb{R}$. So we setting up the following diagram of functions:



(3.2)

Now we define $(f \times g) : [0, 1] \to \mathbb{R}^2$ by $(f \times g)(t) = (f(t), g(t))$. The result is that we create a parametric equation in the yz-plane. If we view this parametric equation in 3-dimensions then we not only see its image but we see the natural projections π_1 and π_2 send this image pack to the graphs of f and g in the xy and xz-planes respectively; cf. Figure-3.1.

3.5 Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$ and $g : \mathbb{R} \to \mathbb{R}$ be $g(x) = 1 - x$. Construct explicitly the function $h : \mathbb{R} \to (\mathbb{R} \times \mathbb{R})$ guaranteed by the Universal Mapping Property of Cartesian products. Then graph h in the xyz-plane along with the projections of h to f in the xy-plane and h to g in the xz-plane.

Over time it has become evident that for algebra the properties we need from a Cartesian product can mostly be derived from their Universal Mapping Property. This has lead to a re-definition of sorts in which we allow multiple differing descriptions of direct products to co-exist so long as they all display a Universal Mapping Property. The following exercise demonstrates how this might be encountered.

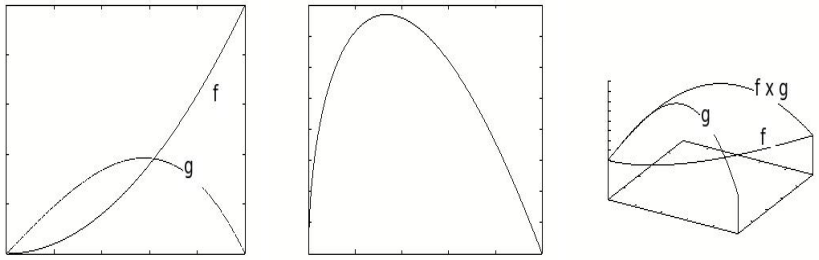


Figure 3.1: The functions f and g; then $f \times g$ drawn parametrically; then f, g, and $f \times g$ drawn together to show how the projections relate to $f \times g$.

3.6 Fix a family $\{A_1, A_2, A_3\}$ of sets. The usual construction $(A_1 \times A_2) \times A_3 = \{((a_1, a_2), a_3) : a_1 \in A_1, a_2 \in A_2, a_3 \in A_3\}$ has projections $\pi_i : (A_1 \times A_2) \times A_3 \to A_i$, $i \in \{1, 2, 3\}$, where $\pi_i((a_1, a_2), a_3) = a_i$. Show that if $\{f_i : T \to A_i : i \in \{1, 2, 3\}\}$ is a family of functions, then there is a unique function $f : T \to (A_1 \times A_2) \times A_3$ such that for every $i \in \{1, 2, 3\}$,

$$\pi_i(f(t)) = f_i(t) \qquad (\forall t \in T, \forall i \in I).$$

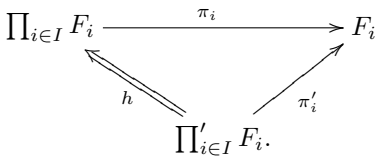
Would this be true of $A_1 \times (A_2 \times A_3)$?

We are prepared to treat products as equivalent if they both have a Universal Mapping Property for Cartesian Products. Yet, we must be careful not to re-introduce the unwanted behavior we had before, such as making $\mathbb{R} \times \mathbb{R}$ the same as $(0, 1) \times (0, 1)$. The following theorem makes the meaning of "same" precise.

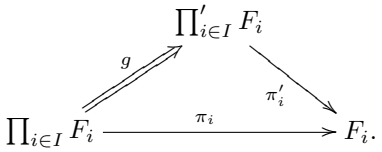
Theorem 3.2.7. *Fix a family $\mathcal{F} = \{F_i : i \in I\}$. Suppose $\langle \prod \mathcal{F}, \pi_i : \prod \mathcal{F} \to F_i \rangle$ and $\langle \prod' \mathcal{F}, \pi'_i : \prod \mathcal{F} \to F_i \rangle$ are two sets with projections such that both behave as Cartesian products in that they both satisfy the Universal Mapping Property for Cartesian Products (i.e. for every family $\{f_i : T \to F_i : i \in I\}$ there are unique functions $f : T \to \prod \mathcal{F}$ and $f' : T \to \prod' \mathcal{F}$ such that for all i, $\pi_i \circ f = f_i$ and $\pi'_i \circ f' = f_i$). It follows that there is a unique invertible function $h : \prod \mathcal{F} \to \prod' \mathcal{F}$ such that for every $i \in I$, $\pi'_i \circ h = \pi_i$ and $\pi_i \circ h^{-1} = \pi_i$.*

The statement of Theorem-3.2.7 is long and difficult to process. However, the concept it conveys is that that two different constructions of a product which has the UMP4CP will be essentially the same. For example, we could let $\mathcal{F} = \{A_1, A_2, A_3\}$ and let $\prod \mathcal{F} = A_1 \times A_2 \times A_3$ be the Cartesian product we defined above so that $\pi_i(a_1, a_2, a_3) = a_i$. Then for the second product we can use $(A_1 \times A_2) \times A_3$ with the projections $\pi'_i((a_1, a_2), a_3) = a_i$. These are different sets and so different products. Yet both have the Universal Mapping Property for Cartesian Products and so the theorem asserts that they are not in-fact that different.

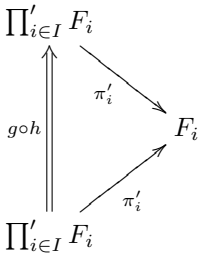
Proof. Using the functions π_i and π'_i we can setup the information we need to apply the Universal Mapping Property for Cartesian Products. Specifically, we let $T = \prod' \mathcal{F}$ in the diagram below and from the UMP4CP applied to $\prod \mathcal{F}$, we obtain a unique function h making the diagram commute, that is, making $\pi_i \circ h = \pi'_i$ for each $i \in I$.



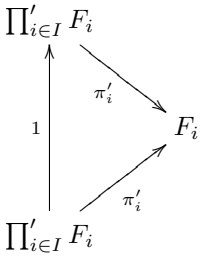
Next we reverse the roles of π_i and π'_i and use the UMP4CP applied to $\prod' \mathcal{F}$ (instead of $\prod \mathcal{F}$ as we did before). This gives us a unique function $g : \prod \mathcal{F} \to \prod' \mathcal{F}$ such that $\pi'_i \circ g = \pi_i$, for each $i \in I$. We capture this in the following diagram.



Now we consider these two diagrams together and by composing g and h we arrive at the diagram:



Notice that $g \circ h$ is the unique function to fit that diagram, i.e. to have $\pi'_i \circ (g \circ h) = \pi'_i$, because we built g and h uniquely to have that property. However, there is another function which works here also, namely the identity function 1 in $\prod'_{i\in I} F_i$ also has $\pi'_i \circ 1 = \pi'_i$.



Since $g \circ h$ is the unique solution but we find 1 also works we are forced to agree that $g \circ h = 1$.

This is half-way towards showing that g and h are inverses. It remains to show that $h \circ g$ is the identity on $\prod \mathcal{F}$. But for that notice we could rewrite the argument above swapping $\prod \mathcal{F}$ with $\prod' \mathcal{F}$ and we would arrive therefore at $h \circ g = 1$.Hence, in fact g and h are inverses and so the two products are uniquely related by an invertible function. □

Because of Theorem-3.2.7 we are now prepared to make the following bold and versatile compromise with Cartesian products. Instead of insisting that our definition in Definition-3.2.3 is the one and only possible meaning of a Cartesian Product, we instead make room for others such as $(A \times B) \times C$ etc. by using the following weaker requirement.

Definition 3.2.8. For a family $\mathcal{F}$, a *Cartesian Product* for $\mathcal{F}$ is a set $P(\mathcal{F})$ together with a family of functions $\{\pi_F : P(\mathcal{F}) \to F : F \in \mathcal{F}\}$ such that whenever $\{f_F : T \to F : F \in \mathcal{F}\}$ is a family of functions, there is a unique function $f : T \to P(\mathcal{F})$ such that $\pi_F \circ f = f_F$, for each $F \in \mathcal{F}$.

This definition should remind you of how we defined Free algebras in Definition-3.1.2 in that the definition is not a specific set but instead it is Universal Mapping Property about the sets and their functions.

3.7 True or False? Using the new definition of Cartesian products, is it true that $\mathbb{R} \times \mathbb{R}$ is the same as $(0,1) \times (0,1)$?

3.2.2 Direct Products

Following on the successful creation of Cartesian products for arbitrary families of sets we do the same for families of sets with operations. There is one sensible restriction. If we have sets with operations that we whish to combine to make one larger sets with operations it makes sense that the initial sets should have the same signature for their operations. That is, we wont take Cartesian products of $\langle S, +, \cdot, 0\rangle$ with $\langle T, \{t_1, \ldots, t_{11}\}\rangle$ as the resulting set $S \times T$ will have no obvious operations. So now suppose $\langle S, \{x_1, \ldots, x_n\}\rangle$ and $\langle T, [t_1, \ldots, t_n]\rangle$ are sets with n-ary operations. Immediately we see how to equip $S \times T$ with an n-ary operation as follows:

$$/(s_1, t_1), \ldots, (s_n, t_n)/ = (\{s_1, \ldots, s_n\}, [t_1, \ldots, t_n]). \tag{3.3}$$

To be very concrete: the set $\mathbb{Q} \times (\mathbb{Z}/12\mathbb{Z})$ has the binary operation of addition defined as follows:

$$(a/b, x + 12\mathbb{Z}) + (c/d, y + 12\mathbb{Z}) = \left(\frac{ad + bc}{bd}, (x + y) + 12\mathbb{Z}\right). \tag{3.4}$$

Of course we could use other binary operation on these two sets instead of addition, for example creating:

$$(a/b, x + 12\mathbb{Z}) + (c/d, y + 12\mathbb{Z}) = \left(\frac{ac}{bd}, (x + y) + 12\mathbb{Z}\right). \tag{3.5}$$

This underscores how important it is to have previously agree on the operations implied which is usually easiest by selecting a common signature for algebra we seek to study.

Now to establish an operation on larger Cartesian products we much ensure that the definition is reliable. This is achieved with the following technical result but which essentially says the same thing we have just demonstrated above.

Lemma 3.2.9. *There is a unique invertible function* $h : \left(\prod_{F\in\mathcal{F}} F\right)^n \to \prod_{F\in\mathcal{F}} F^n$ *such that for all* $F \in \mathcal{F}$, $\pi_{F^n} \circ h = (\pi_F)^n$ *where* $(\pi_F)^n = \prod_{i=1}^n \pi_F$.

3.8 Prove Lemma-3.2.9. [Hint: use the Universal Mapping Property of Cartesian products.]

Theorem 3.2.10. *If* $\mathcal{F}$ *is a family of sets such that each* $F \in \mathcal{F}$ *has an* n*-ary operation* $\{x_1, \ldots, x_n\}_F$ *then there is a unique* n*-ary operation* $\{x_1, \ldots, x_n\}$ *on* $\prod \mathcal{F}$ *such that for each* $F \in \mathcal{F}$,

$$\pi_F(\{x_1, \ldots, x_n\}) = \{\pi_F(x_1), \ldots, \pi_F(x_n)\}_F \qquad (\forall x_1, \ldots, x_n \in \prod \mathcal{F}).$$

Proof. This proof asks us to consider three Cartesian products: $\prod_{F\in\mathcal{F}} F^n$, $\left(\prod_{F\in\mathcal{F}} F\right)^n$, and $\prod \mathcal{F}$. First, by Lemma-3.2.9 there is a unique invertible function $h : \left(\prod_{F\in\mathcal{F}} F\right)^n \to \prod_{F\in\mathcal{F}} F^n$ where for all $F \in \mathcal{F}$, $\pi_{F^n} \circ h = \pi_F^n$.

For each $F \in \mathcal{F}$, we denote the operation $\{x_1, \ldots, x_n\}_F$ as a function $f_F : F^n \to F$. Now consider the family $\mathcal{F}^n = \{F^n : F \in \mathcal{F}\}$. If we form the Cartesian product $\prod \mathcal{F}^n = \prod_{F\in\mathcal{F}} F^n$ then we obtain also projection maps $\pi_{F^n} : \prod \mathcal{F}^n \to F^n$. We may also form the Cartesian product $\prod \mathcal{F}$. For each $F \in \mathcal{F}$, we apply the Universal Mapping Property of Cartesian products to the data $\mathcal{F}$

together with the family functions $\{f_F \circ \pi_{F^n} : F \in \mathcal{F}\}$. This determines, for each $F \in \mathcal{F}$, a unique function $g : \prod \mathcal{F}^n \to \prod \mathcal{F}$ such that $\pi_F \circ g = f_F \circ \pi_{F^n}$.

Now define the n-ary operation on $\prod \mathcal{F}$ as follows:

$$\{x_1, \ldots, x_n\} = g(h(x_1, \ldots, x_n)) \qquad (\forall x_1, \ldots, x_n \in \prod \mathcal{F}). \tag{3.6}$$

It follows that for every $F \in \mathcal{F}$,

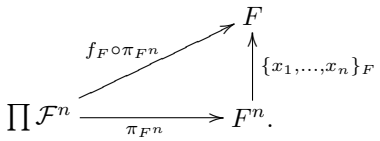
$$\begin{aligned}
\pi_F(\{x_1, \ldots, x_n\}) &= \pi_F(g(h(x_1, \ldots, x_n))) && (3.7)\\
&= (\pi_F \circ g)(h(x_1, \ldots, x_n)) && (3.8)\\
&= (f_F \circ \pi_{F^n})(h(x_1, \ldots, x_n)) && (3.9)\\
&= f_F((\pi_{F^n} \circ h)(x_1, \ldots, x_n)) && (3.10)\\
&= f_F(\pi_F^n(x_1, \ldots, x_n)) && (3.11)\\
&= f_F(\pi_F(x_1), \ldots, \pi_F(x_n)) && (3.12)\\
&= \{\pi_F(x_1), \ldots, \pi_F(x_n)\}_F. && (3.13)
\end{aligned}$$

□

If you have been reading the proof of Theorem-3.2.10 it may take some time to arrange all the behavior. This is not uncommon of such abstract proofs. A useful device in tracking the process is to draw a picture along side which illustrates the various functions as they are introduced. This is a dynamic process so it is difficult to include in a text. The process might look like the following sequence of diagrams.

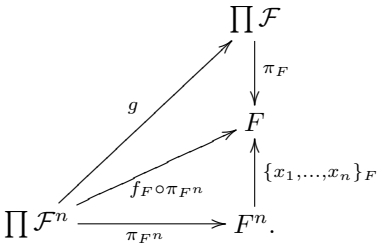
We begin with some initial functions.

(3.14)



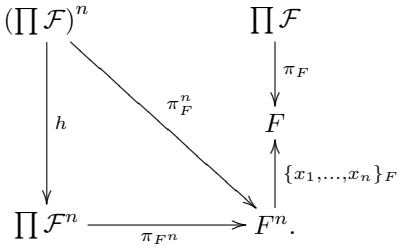
We use these to create the functions $\{f_F \circ \pi_{F^n} : F \in \mathcal{F}\}$ – which does *not* depend on a Universal Mapping Property. Next using the Universal Mapping Property on $\prod \mathcal{F}$ to produce the function g induced from $\{f_F \circ \pi_{F^n} : F \in \mathcal{F}\}$. So now we have:

(3.15)

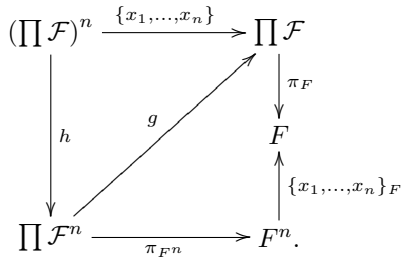


Using Lemma-3.2.9 we obtain a function h which fits into the diagram:

(3.16)

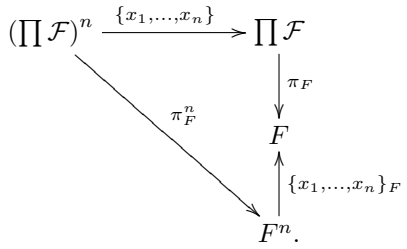


Using h we define $\{x_1, \ldots, x_n\}$ as $g \circ h$ as seen in the diagram.



(3.17)

Now stand back and replace the functions which remain relevant and we see our final diagram visually illustrates our effort is correct.



(3.18)

That is to say, $\pi_F \circ \{x_1, \ldots, x_n\} = \{x_1, \ldots, x_n\}_F \circ \pi_F^n$.

Definition 3.2.11. Fix a variety $\mathfrak{V}$ with signature σ. The *direct product* of a family $\mathcal{F}$ of members of $\mathfrak{V}$ is the Cartesian product $\prod \mathcal{F}$ equipped with operation of each type in σ according to the construction in Theorem-3.2.10.

3.9 Let $\mathcal{F}$ be a family of $\{\cdot, 1\}$-algebras. Show that $\mathcal{F} \subseteq \mathfrak{V}(x \cdot 1 = x = 1 \cdot x)$ then $\prod \mathcal{F} \in \mathfrak{V}(x \cdot 1 = x = 1 \cdot x)$. In particular, give the identity of $\prod \mathcal{F}$.

3.10 True or False? If $\mathcal{F}$ is a family of $\{\cdot, 1\}$-algebras where at least one member does not have an identity, is it possible for $\prod \mathcal{F}$ to have an identity? Prove your claim.

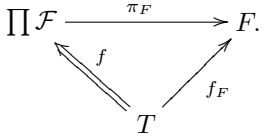
Lemma 3.2.12. *If $\mathcal{F}$ is a family of σ-algebras then for every $F \in \mathcal{F}$, the projection function $\pi_F : \prod \mathcal{F} \to F$ is a σ-homomorphism.*

3.11 Prove Lemma-3.2.12 in the case where σ has one n-ary operation. (Generalizing to arbitrary signature follows by induction – you do not need to prove that.)

Theorem 3.2.13 (Universal Mapping Property for Direct Products)**.** *If $\mathcal{F}$ is a family of σ-algebra, T is a σ-algebra, and $\{f_F : T \to F : F \in \mathcal{F}\}$ is a family of σ-homomorphisms then there is a unique σ-homomorphism $f : T \to \prod \mathcal{F}$ such that*

$$(\forall F \in \mathcal{F}, \forall t \in T), \quad \pi_F(f(t)) = f_F(t).$$

As a diagram of functions this means:



(We use the double-bar function to indicate that homomorphism is unique with respect to all others in the diagram.) The homomorphism f is often denoted $\prod_{F \in \mathcal{F}} f_F$.

Proof. First we know that every σ-homomorphism is a function. Therefore, the Universal Mapping Property of Cartesian Products provides a unique function $f : T \to \prod \mathcal{F}$ such that for each $F \in \mathcal{F}$, $\pi_F \circ f = f_F$. We must show f is a homomorphism. We leave this as an exercise. □

3.12 Show that the f in Theorem-3.2.13 is a homomorphism.

We close by noticing that varieties are closed to direct products. This means that we have an alternative method to construct examples of members in a variety.

Theorem 3.2.14. *Fix a variety $\mathfrak{V}$ of σ-algebras. If $\mathcal{F} \subset \mathfrak{V}$ then $\prod \mathcal{F} \in \mathfrak{V}$.*

3.13 Show directly that if G and H are monoids then $G \times H$ is a monoid.

3.14 Prove that if a variety has one member of size greater than 1 then the variety is infinite, i.e. the variety has an infinite number of members. [Hint: consider direct products.]

3.15 True or False? We say a group G is finite if the size of the set G is finite. Is the class of finite groups a variety?

For the next exercise we need to settle some standard notation. When we talk about abelian groups we typically write use the signature $\{+, -, 0\}$ so that we are aware the addition is commutative. The direct product of abelian groups A and B is therefore denoted as $A \oplus B$ and for a general family $\mathcal{A}$ of abelian groups by $\bigoplus \mathcal{A} = \bigoplus_{A \in \mathcal{A}} A$.

3.16 True or False? An abelian group A is said to be a *torsion* group if for every element $a \in A$, there is a positive integer n such that $na = \overbrace{a + \cdots + a}^{n} = 0$. Does the class of all torsion abelian groups make a variety? [Hint: consider $(\mathbb{Z}/2\mathbb{Z}) \oplus (\mathbb{Z}/3\mathbb{Z}) \oplus (\mathbb{Z}/5\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p\mathbb{Z}) \oplus \cdots)$ where p runs through every prime.]

Recall in Section 1.6.2 that we introduced fields as commutative rings in which every nonzero element has an inverse. For example, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields but $\mathbb{Z}$ is not a field. We also remarked that the definition of a field failed to specify a variety because in a variety every element of the algebra must satisfy every law. However, with fields we do not require (and indeed cannot require) that the zero have an inverse. So this suggests that perhaps the class of all fields does not make a variety. Yet, we could not prove that at the time. Now we can. We have just proved that varieties are closed to direct products. So if we suppose that the class of fields is a variety then the direct product $\mathbb{Q} \oplus \mathbb{Q}$ must be a fields, because $\mathbb{Q}$ is a field. However, that is not the case because $(1, 0) \in \mathbb{Q} \oplus \mathbb{Q}$ has no inverse. Therefore the class of fields is not closed to direct products and so the class of fields is not a variety.

3.17 Prove that in $\mathbb{Q} \oplus \mathbb{Q}$, $(1, 0)$ has no inverse.

Chapter 4

Lattices, Representations, and Monomorphisms

In Chapter 2 we saw the power of relating equivalence, partitions, and surjections and then extended this interplay to algebra using congruence, quotients, and epimorphisms. It probably did not escape notice that many functions are not surjective so a suspicion may have arrisen asking for the roles other functions, such as one-to-one (injective) functions. This chapter will explore a parallel relationship involving injective functions. However, we caution the reader not to expect a parallel development to that in Chapter **??**. For example, we will not develop any sort of "co-equivalence relation". Though there are connections which we explore later, the concepts of this chapter are somewhat unrelated to the reasoning in the previous chapter.

4.1 Partial ordering

In the integers, rationals, and reals we have the familiar and important tool known as *ordering*. Specifically, every number x and y in these sets can be compared, say, $x \leq y$ or $y \leq x$. This ability is important but somewhat special. There are other properties of $\leq$ which seem to occur with greater frequency. We collect these under the name of a *partial ordering*.

Definition 4.1.1. A relation R on a set S is a *partial ordering* if it satisfies

Reflexive for all $s \in S$, sRs.

Anti-symmetric for all $s, t \in S$, if sRt and tRs then $s = t$.

Transtive for all $s, t, u \in S$, if sRt and tRu then sRu.

4.2 Injectivity, subsets, and lattices

Once again, we begin with the implications from Set Theory and gradually adapt these to algebra. Recall that a relation from A to B is simply a subset $R \subseteq A \times B$. For every relation we can consider its dual $R^{-1} \subseteq B \times A$ defined by

$$R^{-1} = \{(b, a) : (a, b) \in R\}.$$

For example, if $R = \left\{(x, y) : x^2 + \frac{y^2}{4} = 1\right\} \subseteq \mathbb{R} \times \mathbb{R}$ then the graph of our relation R is an ellipse with x-radius 1 and y-radius 2. So R^{-1} is an ellipse with x-radius 2 and y-radius 1 (because we simply swap the coordinates of every point in R). The notation R^{-1} is suggestive of inverse but for now we should treat it just as notation.

Now suppose we consider functions. There are two properties that establish a relation $R \subseteq A \times B$ as a function $f : A \to B$, where $f(a) = b$ if, and only if, $(a, b) \in R$.

(i) For every $a \in A$ there is a $b \in B$ such that $(a, b) \in R$. In functional vernacular we say A is the domain of f and B is the codomain.

(ii) For every $a \in A$, if $b, b' \in B$ such that $(a, b), (a, b') \in R$ then $b = b'$. Again in functional speak we say that f is well-defined.

If we think visually, property (i) says that the graph of f has no 'holes', no 'asymptotes', i.e. no place in A which is not given a point on the graph. Property (ii) is often described as the "vertical line test", i.e. that for every input $a \in A$, there is only one output.

Now, if $R \subseteq A \times B$ establishes a function $f : A \to B$ as above, we can temporarily forget that R is a function and just consider its dual R^{-1}. What we often find is that R^{op} does not determine a function $f^{-1} : B \to A$. As with R^{-1}, f^{-1} anticipates that we eventually wish to relate f to an inverse function. However, as there are two conditions a function must satisfy, there are also two ways in which R^{-1}, and therefore f^{-1}, can fail to be a function.

If R^{-1} fails property (i) then we are saying that for some $b \in B$ there is no $a \in A$ such that $(b, a) \in R^{-1}$, and so for all $a \in A$, $(a, b) \notin R$. This is simply saying that our function $f : A \to B$ is not surjective: there is a $b \in B$ with no $a \in A$ such that $f(a) = b$.

On the other-hand, if R^{-1} fails property (ii) then we notice that for some value $b \in B$ there exist $a, a' \in A$ such that $a \neq a'$ yet (a, b) and (a', b) both lie in R. Translating this to the f notation we are saying that form some $a, a' \in A$, $f(a) = f(a')$ yet $a \neq a'$. This suggests that we invent a description of functions that avoid that problem. In fact this is the well-known property of 'one-to-one' functions.

Definition 4.2.1. A partial function $f : A \to B$ is *injective* if

$$(\forall a, a; \in A) \qquad f(a) = f(a') \Rightarrow a = a'.$$

We also say that f is *one-to-one* or an *injection*.

Theorem 4.2.2. *(i) If $f : A \to B$ is an injective function then $f^{-1} : B \to A$ is a partial function.*

(ii) If $f : A \to B$ is a surjective function, then $f^{-1} : B \to A$ has domain A.

(iii) If $f : A \to B$ is bijective then f^{-1} is a function.

4.1 Prove Theorem-4.2.2.

4.3 Group Lattices

Proposition 4.3.1. *If H is a subgroup of G if and only if the relation $a \equiv_H b$ defined as $ab^{-1} \in H$ is a right congruence relation.*

Theorem 4.3.2 (Lagrange)**.** *Given a G and a subgroup $H \leq G$ then $|G| = [G : H]|H|$.*

Proof. The main idea is that H partitions G into cosets. Cosets are all in one-to-one correspondance with H. The principle reason for the bijections is the existance of inverses.

From Proposition-4.3.1 we know H partitions G into equivalence classes. Indeed, $a \equiv_H b$ implies $ab^{-1} \in H$ so for all $h \in H$, $a \equiv_H ha$ as $aa^{-1}h \in H$. Consequently $[a] = Ha$. Therefore G is partitioned into cosets $Ha_1, \ldots, Ha_n$ where by definition $n = [G : H]$.

Now given any $a, b \in G$, $f_{a,b} : Ha \to Hb$ via $f(g) = ga^{-1}b$. Clearly the map is well-defined.

$$f_{b,a}(f_{a,b}(g)) = f_{b,a}(ga^{-1}b) = ga^{-1}bb^{-1}a = g = 1(g).$$

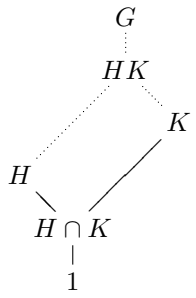
Therefore $f_{a,b}$ is invertible and so also a bijection. Therefore $|Ha| = |Hb|$.

G is the disjoint union of $[G : H]$ many cosets each of size $|H|$, so $|G| = [G : H]|H|$. □

Corollary 4.3.3 (Tower Law)**.** *Given $K \leq H \leq G$ we have*

$$[G : K] = [G : H][H : K]$$

Theorem 4.3.4 (Parallelogram Law)**.** *Given $H, K \leq G$, recall HK is the complex. Opposite sides of the parallelogram are congruent.*



That is:

$$[HK : H] = [K : H \cap K].$$

In particular

$$[G : H] \geq [K : H \cap K].$$

Proof. The essential idea is that $f : K/H\cap K \to HK/K$, given by $f(gH\cap K) = gH$, is a bijection.

Given $gH\cap K = kH\cap K$ we have $k^{-1}gH\cap K = H\cap K$ so in fact $k^{-1}g \in H\cap K$. Consequently,

$$f(k^{-1}gH \cap K) = k^{-1}gH.$$

□

Lemma 4.3.5. *$H, K \leq G$, then HK/K and KH/K are naturally equivalent as left or right cosets.*

4.4 Lattices of Subgroups

Definition 4.4.1. Normal A subgroup H is normal in G if for every $\alpha \in \mathrm{Inn}(G)$ $\alpha(H) \leq H$. [As α is invertible we prove $\alpha(H) = H$.]

Characteristic A subgroup H is characteristic in G if for every $\alpha \in \mathrm{Aut}(G)$, $\alpha(H) \leq H$. [As α is invertible we prove $\alpha(H) = H$.]

S-Invariant A subgroup H is *S-invariant* in G if for every $\alpha \in S \leq \mathrm{End}(G)$, $\alpha(H) \leq H$.

Fully Invariant A subgroup H is *fully invariant* in G if for every $\alpha \in \mathrm{End}(G)$, $\alpha(H) \leq H$.

Theorem 4.4.2. *All S-invariant subgroups of G from a complete lattice.*

Proof. Let $\{H_i \mid i \in I\}$ be a family of subgroups of a group G. Then we define the intersection to be setwise intersection. Likewise, we take the join to be the intersection of all subgroups which contain every subgroup in the family.

Given a family $\{H_i \mid i \in I\}$ of characteristic subgroups, if we let $\alpha \in S$ and $x \in \bigcap_{i\in I} H_i$ then we find $\alpha(x) \in H_i$ for all i since each H_i is characteristic. Indeed then we have $\alpha(x) \in \bigcap_{i\in I} H_i$ proving $\bigcap_{i\in I} H_i$ is characteristic in G.

Also given any that the characteristic subgroups are normal, the join is the product; therefore, every element is of the form $x_{i_1} \cdots x_{i_j}$ where $x_{i_k} \in H_{i_k}$ and consequently

$$\alpha(x_{i_1} \cdots x_{i_j}) \in H_{i_1} \cdots H_{i_j} \leq \Pi_{i \in I} H_i.$$

Therefore, $\Pi_{i \in I} H_i$ is characteristic. □

Proposition 4.4.3. *Given any group G, $\mathrm{End}(G)$ acts on the subgroup lattice of G and is order preserving. In particular $\mathrm{Aut}(G)$ acts as a group on G.*

Proof. For any $H \leq G$, $1(H) = H$. Likewise taking any $\alpha, \beta \in \mathrm{End}(G)$ we find $\alpha(\beta(H)) = \alpha\beta(H)$. □

In general, the action need not be faithful, and it is transitive only when $G = 1$ – by the order preservation.

The characteristic subgroups are the fixed points of the action of $\mathrm{Aut}(G)$ on $\mathcal{L}(\mathcal{G})$.

By the Sylow theorems we know that all Sylow subgroups are conjugate. In particular, $\mathrm{Aut}(G)$ acts transitively on the Sylow-p-subgroups but the only necessary members are in $\mathrm{Inn}(G)$. So if $\mathrm{Aut}(G) \neq \mathrm{Inn}(G)$ then $\mathrm{Aut}(G)$ contains a non-trivial kernel.

4.5 Group Actions

Definition 4.5.1. Given a group G and a set X, a group action is a map $a : X \times G \to X$ where

1. $a(x, 1) = x$ for all $x \in X$
2. $a(a(x, g), h) = a(x, gh)$ for all $x \in X$ and $g, h \in G$.

We express the action as

$$a(g, x) = x^g.$$

The set X is called a G-set. The following sets all are natural consequences: let $x \in X$ and $g \in G$,

Orbits $x^G = Gx = \{y \in X \mid x^g = y, \text{ for some } g \in G\}$,

Fixed Points $Fix(g) = X_g = \{x \in X \mid x^g = x\}$,

Fixed Points $Fix(G) = X_G = \{x \in X \mid x^g = x, \text{ for all } g \in G\}$,

Index $[X : G]$ the number of orbits in G.

Stabilizer $\mathrm{Stab}_G(x) = G_x = \{g \in G \mid x^g = x\}$,

Kernel $\ker\ a = \{g \in G \mid x^g = x, \text{ for all } x \in X\}$.

Proposition 4.5.2. *The oribits partition X. Stablizers are all subgroups of G and the kernel is the intersection of the stabilizers; moreover, the kernel is a normal subgroup. Finally*

$$[G : \mathrm{Stab}_G(x)] = |x^G|.$$

Proposition 4.5.3 (Cayley)**.** *The category of all group actions on a set X is isomorphic to the category of homomorphisms into $Sym(X)$. In particular, given a group action $\langle G, X, a \rangle$, $\Gamma_a : G \to Sym(X)$ is given by*

$$\Gamma_a(g)(x) = x^g.$$

Also, $ker\ \Gamma_a = ker\ a$.

Definition 4.5.4. A group action is called:

Faithful if $ker\ a = 1$,

Free if $Fix(G) = \varnothing$,

Transitive if $x^G = X$ for any $x \in X$,

Regular if $\mathrm{Stab}_G(x) = 1$ for all $x \in X$. In particular all regular actions are faithful, free, and transitive,

Primitive if it is transitive and $\mathrm{Stab}_G(x)$ is a maximal subgroup of G for any $x \in X$.

Nearly without exception, every group action should be considered as faithful. This is because every group action $\langle G, X, a\rangle$ can be replaced by $\langle G/Ker\ a, X, a\rangle$ with no substintive change to the theory and results.

Free actions are typically of interest to topologist. Group thoery makes use of actions of groups on themselves and each other, and consequently elements that are central or identity tend to be fixed points and so free actions are not that important. Finally, the fact that primitivity is well-defined despite the choice of x is due to that following result.

Proposition 4.5.5. *Given $g \in G$ and $x \in X$*

$$g^{-1}\,\mathrm{Stab}_G(x)g = \mathrm{Stab}_G(x^g)$$

in particular, if G acts transitively, then all stabilizers are conjugate and indeed isomorphic.

Proof. Given any $h \in \mathrm{Stab}_G(x)$ it follows $x^h = x$ and so

$$(x^g)^{g^{-1}hg} = x^{hg} = x^g.$$

Hence $g^{-1}hg \in \mathrm{Stab}_G(x^g)$. □

Proposition 4.5.6 (Class Equation)**.** *Given a finite group G and a finite set G-set X, then*

$$|X| = |Fix(G)| + \sum_{x \in T} [G : \mathrm{Stab}_G(x)]$$

where T is a transversal of all orbits which are not singletons.

Corollary 4.5.7. *Given any finite group G,*

$$|G| = |Z(G)| + \sum_{g \in T} [G : C_G(g)].$$

Lemma 4.5.8 (not-Burnside's Lemma)**.** *Given a finite group G and a finite G-set X, it follows*

$$[X : G] = \frac{1}{|G|} \sum_{g \in G} |Fix(g)|$$

Definition 4.5.9. Given a fixed group G and two G-sets Γ and Ω we define a G-map $f : \Gamma \to \Omega$ as any function of sets where $f(\gamma^g) = f(\gamma)^g$ for all $\gamma \in \Gamma$ and $g \in G$.

We say two G-sets Γ and Ω are G-isomorphic, and write $\Gamma \simeq_G \Omega$, if there is a pair of G-maps $f : \Gamma \to \Omega$ and $g : \Omega \to \Gamma$ such that $fg = 1_\Omega$ and $gf = 1_\Gamma$.

It is clear that composition of G-maps is again a G-map so we can consider the category of all (right) G-sets, denoted $\mathbf{Set}_G$ with G-maps as morphisms.[1]

Proposition 4.5.10 (Cayley)**.** *Let $\mathcal{G}$ denote the category of all groups. The following categories are all naturally isomorphic:*

(i) $\boldsymbol{Set}_{\mathcal{G}}$

(ii) ${}_{\mathcal{G}}\boldsymbol{Set}$

(iii) ***PermGroups*** *– the full subcategory (of $\mathcal{G}$) of all permutation groups.*

We also have a version of the isomorphism theorems which is as transparent as for groups.

Lemma 4.5.11. *A bijective G-map is a G-isomorphism.*

Proof. Given a G-map $f : \Omega \to \Gamma$ which is bijective, then

$$\omega^g = f(f^{-1}(\omega))^g = f(f^{-1}(\omega)^g);$$

hence,

$$f^{-1}(\omega^g) = f^{-1}(f(f^{-1}(\omega)^g)) = f^{-1}(\omega)^g.$$

Thus f^{-1} is a G-map so f is a G-isomorphism. □

Proposition 4.5.12. *Given a transitive G-set Ω, then there is a canonical G-isomorphism $R : \Omega \to G/\operatorname{Stab}_G(\omega)$ for any $\omega \in G$. Moreover, given any $g \in G$ we find*

$$\operatorname{Stab}_G(\omega^g) = g^{-1}\operatorname{Stab}_G(\omega)g = \operatorname{Stab}_G(\omega)^g$$

and

$$G/\operatorname{Stab}_G(\omega) \simeq_G \Omega \simeq_G G/\operatorname{Stab}_G(\omega^g).$$

Proof. Define $R : \Omega \to G/\operatorname{Stab}_G(\omega)$ as $R(\omega^g) = \operatorname{Stab}_G(\omega)g$.

Since G acts transitively on Ω then for every $\gamma \in \Omega$ there exists a $g \in G$ such that $\gamma = \omega^g$. Moreover, given two such $g, g' \in G$ where $\omega^g = \omega^{g'}$ it follows $\omega = \omega^{g'g^{-1}}$ so that $g'g^{-1} \in \operatorname{Stab}_G(\omega)$. Since $\operatorname{Stab}_G(\omega)$ is a subgroup it follows $g \equiv g' \pmod{\operatorname{Stab}_G(\omega)}$ proving R is well-defined.

Next $\operatorname{Stab}_G(\omega)g = \operatorname{Stab}_G(\omega)g'$ if and only if $g \equiv g' \pmod{\operatorname{Stab}_G(\omega)}$ so that $\omega^g = \omega^{g'}$ proving R is injective. Finally, given any $\operatorname{Stab}_G(\omega)g$ clearly $\omega^g \in \Omega$ so that R is surjective.

Given $g, g' \in G$,

$$R((\omega^g)^{g'}) = R(\omega^{gg'}) = \operatorname{Stab}_G(\omega)gg' = R(\omega^g)g'.$$

Therefore R is a bijective G-maps so it is a G-isomorphism.

Given any $g \in G$ and $h \in \operatorname{Stab}_G(\omega)$ then

$$(\omega^g)^{h^g} = (\omega^g)^{g^{-1}hg} = (\omega^h)^g = \omega^g.$$

Thus $h^g \in \operatorname{Stab}_G(\omega^g)$ so that $\operatorname{Stab}_G(\omega)^g \leq \operatorname{Stab}_G(\omega^g)$. By the symmetric argument $\operatorname{Stab}_G(\omega)^g = \operatorname{Stab}_G(\omega^g)$.

Finally, given a that $\omega^G = \Omega = (\omega^g)^G$ it follows R_ω and R_{ω^g} are both G-isomorphisms so that

$$G/\operatorname{Stab}_G(\omega) \simeq_G \Omega \simeq_G G/\operatorname{Stab}_G(\omega^g).$$

□

[1] Left acting G-sets would be denoted ${}_G\mathbf{Set}$. If G is abelian then every G-set is both left and right in the natural way.

Proposition 4.5.13. *Given $H, K \leq G$ then KH, KH/K, and $H/H \cap K$ are natural H-sets and furthermore,*

$$KH/K \simeq_H H/H \cap K.$$

[Assuming action on the right. For an action on the left: $HK/K \simeq_H H/H \cap K$.]

Proof. Given any $kh \in KH$ and $h' \in H$ define $kh^{h'} = khh'$. Clearly then $khh' \in KH$ as $hh' \in H$. The axioms for the action follow equally painless: $kh^1 = kh1 = kh$, $(kh^a)^b = khab = kh^{ab}$. We see the argument is identical when we consider KH/K and the action $Kh^{h'} = Khh'$. Finally, $H \cap Kh^{h'} = H \cap Khh'$ completes the listing of the natural actions.

For the isomorphism, it suffices to build a bijective G-map. Define $f : H/H \cap K \to KH/K$ by $H \cap Kh \to Kh$. To see that f is well-defined take $h \equiv h' \pmod{H \cap K}$, that is, $h^{-1}h' \in H \cap K$ so that $h^{-1}h' \in K$ proving $h \equiv h' \pmod{K}$. Since $h, h' \in H$ it follows $h^{-1}h' \in H$ so the steps are reversable proving also that f is injective. Finally it is clear the f is surjective so we need only prove that f is an H-map. Given again any $h, h' \in H$ we have

$$f(H \cap Khh') = Khh' = f(H \cap Kh)h'.$$

□

Definition 4.5.14. Given a G-set Ω we define the G-topology of Ω as the topology generated by the orbits of Ω under the action of G.

Proposition 4.5.15. *Every G-map is G-continuous.*

Proof. Orbits partition Ω so pre-images of basic open sets are open sets. □

Proposition 4.5.16. *A G-set is transitive G-set if and only if it is connected.*

4.6 Faithful Actions

An action of G on a set Ω is equivalent to asking for a homomorphism $\varphi : G"->"Sym(\Omega)$. The kernel of this homomorphism is the *kernel of the action.* An action is *faithful* when the kernel is trivial. We also speak of a permutation representation for G when given such a homomorphism.

By Cayley's theorem we know every group has a transitive faithful permutation representation, namely, the regular representation. However this is on a set equal the size of the group which may be too large for meaningful computations.

Not all groups can be faithfully represented on small sets.

Proposition 4.6.1. *The kernel of an action is the intersection of the stabilizers of the action. When the action is transitive this is equivalent to the core of any stabilizer.*

Recall that a p-group G is extraspecial if $Z(G) = G' = \Phi(G) \cong C_p$.
Example: An extraspecial p-group G of order p^{2m+1} has no transitive faithful permutation representation of degree less than 2^{m+1}. □

Proof. Let H be a subgroup of G of index $[G : H] \leq 2^m$ which we wish to have as a stabilizer of a transitive action. If H is normal in G then it cannot be a stabilizer of a faithful transitive action (it is its own kernel.) As such we assume that H is not normal in G and the core of H is trivial.

As G/G' is abelian, any subgroup H of G where $G' \leq H \leq G$, is immediately normal in G (H/G' is normal in G/G'.) Thus $H' = H \cap G' = 1$ so H is abelian. Likewise $G^p \leq \Phi(G)$ in any p-group so $H^p \leq \Phi(G)$. But $\Phi(G)$ is a minimal subgroup so $H^p = 1$ or $\Phi(G)$. As $\Phi(G) = G'$ we know $H^p = 1$. Therefore H is elementary abelian.

Now we use geometry to complete the proof. First we notice that HZ/Z is a totally isotropic/totally singular subspace as $H \cap Z = 1$. (For any $a, b \in H$, $[a, b] = 1$ or $\varphi(a) = a^2 = 1$ in the $p = 2$ case.) Thus HZ/Z is contained in a maximal totally isotropic/singular subspace. However maximal totally istropic/singular subspace have dimension m. Thus HZ/Z is too large to be contained in such a subspace.

Hence H must be normal and contain Z. Hence every large stablizer has a kernel. □

Given that there are not transitive faithful embeddings of an extraspecial group we consider intransitive actions.

Corollary 4.6.2. *There is no faithful permutation representation of an extraspecial group of degree less than p^m.*

Proof. We saw that every large stabilizer contains the center of G. Thus the intersection of all the stabilizers of an intransitive action of small degree still contains the center and thus is not faithful. □

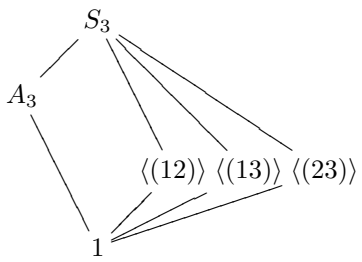
Given a faithful transitive group action of G on Ω, that is, G is embedded in $Sym(\Omega)$, we attach the subgroup structure of G on Ω and visa-versa. In particular we make connections between the lattice of all subsets of Ω and the lattice of all subgroups of G. We begin with a comprehensive example.

4.7 The Action of S_4 on 4 Elements

Let $\Omega = \{1, 2, 3, 4\}$ and $G = Sym(\Omega) = S_4$. The thirty subgroups of S_4 are too many to render in a two dimensional lattice with any decernible properties. However three dimensional models reveal a surprising amount of symmetry as may be expected as the group is the group of symmetries of a cube (in this case the action is in S_8.)

In general when studying a group action the first subgroup to locate is the kernel. Since S_4 acts faithfully we may skip to the stablizer of a point.

Fixing any point in Ω means we are free to permute any of the remaining 3, so we expect to find $Stab_{S_4}(1) \cong S_3$, as it is. Independently we can study the subgroup structure of S_3.



Now we may ask: what is the relationship with the subgroups of $Stab_{S_4}(1)$ and subsets of Ω?

4.7.1 Fixing Subgroups

Definition 4.7.1. Given a permutation group G in $Sym(\Omega)$, and any subset Δ of Ω, we let $Fix_G(\Omega)$

4.7.2 Restricted Actions

Given a subgroup $H \leq G$ and G acting on Ω we can restrict the action to H. The restricted action can be useful to decompose G, say for instance by a Frattini argument. Additionally, normal subgroups provide nice consequences which are obvious when once we prove the following set of theorems.

Lemma 4.7.2.

Theorem 4.7.3.

4.8 Primitive Group Actions

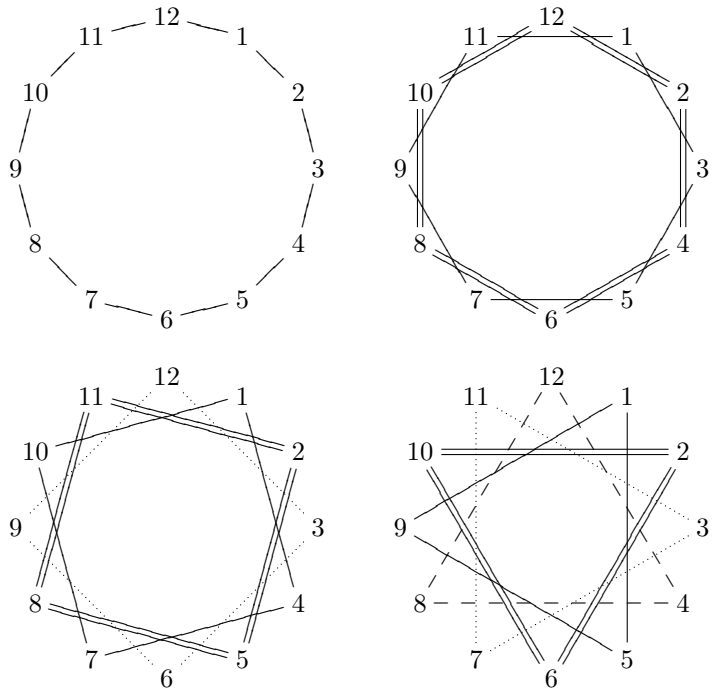
Given a group action of G on a set Ω, we may immediately assume the action is transitive on the orbits. To this end we take all actions on Ω to be transitive unless specified otherwise. A further *divide-and-conquer* process is to decompose Ω into blocks where the action of G is split into the internal action inside each block and the external action which permutes the blocks.

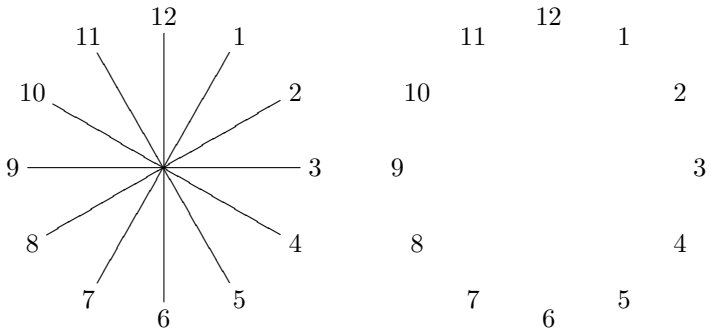
Definition 4.8.1. Given a group action of G on Ω, a **block system** for G is any partition $\mathcal{B}$ of Ω, whose elements we call **blocks**, and where given any block $B \in \mathcal{B}$, $B^g \in \mathcal{B}$.

A block system is called cyclic if there exists a block B such that $\mathcal{B} = \{B^g \mid g \in G\}$.

Because of the existence of inverses, it is clear that in a cyclic block system, all blocks are generators.

Example: The block system for the natural action of D_{24} on the 12 vertices of the dodecagon can be seen in the inscribed geometry.





This gives us the following block systems:

$$1,2,3,4,5,6,7,8,9,10,11,12, \qquad 1,3,5,7,9,11|2,4,6,8,10,12,$$
$$1,4,7,10|2,5,8,11|3,6,9,12, \qquad 1,5,9|2,6,10|3,7,11|4,8,12,$$
$$1,7|2,8|3,9|4,10|5,11|6,12, \qquad 1|2|3|4|5|6|7|8|9|10|11|12.$$

□

Proposition 4.8.2. *A group action is transitive if and only if every block system is cyclic.*

Proof. Let G act on Ω.

Let $\mathcal{B}$ be a block system of a group action of G on Ω. Given any blocks B and C in $\mathcal{B}$, take $b \in B$ and $c \in C$. Since G acts transitively on Ω, there exists a $g \in G$ such that $b^g = c$; hence, $B^g \cap C \neq \varnothing$. However $\mathcal{B}$ is a partition so we conclude $B^g = C$.

For the converse, take the trivial partition of

$$\mathcal{B} = \{\{\omega\} \mid \omega \in \Omega\}.$$

Given any two elements $\omega_1, \omega_2 \in \Omega$, we may take $\{\omega_1\}$ to generate $\mathcal{B}$ so that there exists a $g \in G$ where $\omega_1^g = \omega_2$. Hence the action is trivial. □

Remark 4.8.3. Most authors refer to block systems only on transitive group actions. In such cases, block systems are precisely cyclic block systems, and thus the term block system is used without additional qualifiers.

Once we restrict our attention to transitive gorup actions we find the cyclic nature allows us to return to the group for structure. In particular, we now replace block systems with the any generating block. In order to extract comparisons of block systems we choose generating blocks which each have an element in common.

Proposition 4.8.4. *Given a transitive group action of G on Ω, and $\omega \in \Omega$, then any subset $B \subset \Omega$ containing ω is a **block** if and only if $B^g \cap B \neq \varnothing$ implies $B^g = B$. Such a block we call and ω-block, and the block system they induce is an ω-block system, denoted B^G.*

Proof. Since G acts transitively, Ω is the union of B^g for all $g \in G$. Moreover, the sets B^g are pairwise disjoint by assumption so this is indeed a block system. □

Definition 4.8.5. Given two ω-block A and B, we say $A^G \leq B^G$ if and only if $A \subseteq B$. We also set $A^G \wedge B^G = (A \cap B)^G$ and define $A^G \vee B^G$ as the intersection of all ω-block systems C^G where $A \cup B \subseteq C$.

Proposition 4.8.6. *The ordering on ω-blocks forms a complete lattice. In fact, the lattice is isomoprhic to the quotient lattice of $G/\operatorname{Stab}_G(\omega)$. We call the lattice, the ω-block lattice, or simply the block lattice.*

Proof. We use the natural G-isomorphism $f : \Omega \to G/\operatorname{Stab}_G(\omega)$ of coset action: $f(\omega^g) = \operatorname{Stab}_G(\omega)g$. □

Definition 4.8.7. A transitive group action is **primitive** if it has exactly two block systems – namely the two trivial partitions.

Notice that this definition avoids allowing the trivial action to be considered as primitive.

Corollary 4.8.8. *Given a transitive group action of G on Ω, all the following are equivalent:*

(i) The action is primitive.

(ii) The block lattice is simple.

(iii) For some $\omega \in \Omega$, $\operatorname{Stab}_G(\omega)$ is maximal in G.

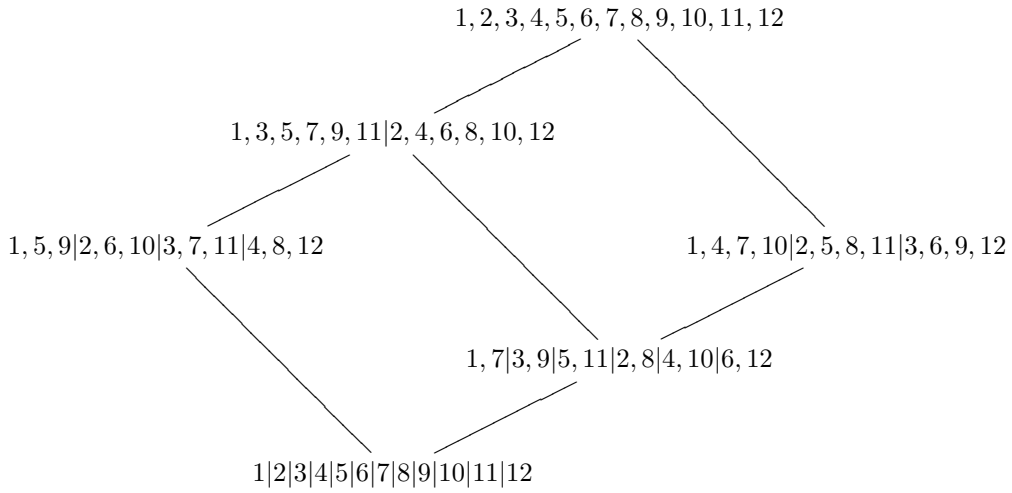
(iv) For all $\omega \in \Omega$, $\operatorname{Stab}_G(\omega)$ is maximal in G.

Proof. A lattice is simple if it has exactly two elements, thus every primitive action produces a simple block lattice. Since the block lattice is order isomorphic to the quotient lattice of subgroup of $G/\operatorname{Stab}_G(\omega)$, it follows $\operatorname{Stab}_G(\omega)$ is maximal in G. Since the choice of ω is arbitrary it is true for all ω. □

Example: Claim: The block lattice of the natural action of D_{2n} on n elements is isomorphic to the subgroup lattice of C_n.

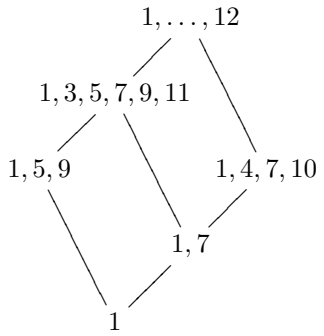
The proof involves induction to show every stabilizer of D_{2n} has quotient lattice isomorphic to C_n. For now an example will suffice with D_{24}.

Originally we would be tempted to write:

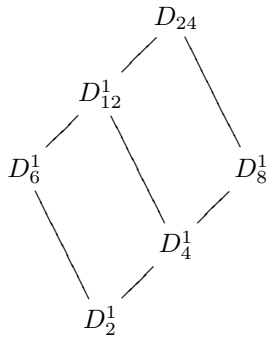


Now using the idea of ω-blocks, we let $\omega = 1$ and write only the block which

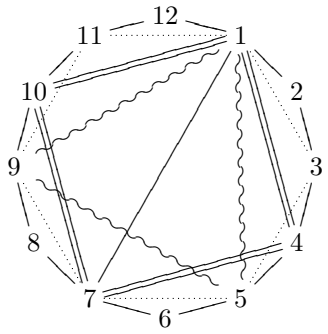
contains ω so we get the simpler listing:



We notice this corresponds to the subgroups of D_{24} as claimed by the theorem:



Notice that geometrically the ω-blocks are all blocks that share a point. We can illustrate this now as:



□

Proposition 4.8.9. *If* $\mathrm{Stab}_G(\omega) \leq N \trianglelefteq G$ *then* $\mathrm{Stab}_G(\omega^g) \leq N$ *for all* $g \in G$.

Proof. Since N is normal it is closed to conjugation; thus,

$$\mathrm{Stab}_G(\omega^g) = \mathrm{Stab}_G(\omega)^g \leq N.$$

□

Definition 4.8.10. A block system is a **normal** block system if it corresponding subgroup in G is normal in G.

The question is whether invariants of group action can detect normality of subgroups. It is known that the block systems of a group action can be determined in polynomial time. Since the degree of a group action is typically logrithmically less than the order of the group, computations that make use of the set structure can be advantageous.

4.9 Classification of Primitive Actions: O'Nan-Scott

Recall that with an abstract action we can consider the action restricted to any subgroup. Suppose that the subgroup is normal, what then might we conclude about the orbits induced?

Lemma 4.9.1. *Given a G-set Ω and a subgroup H of G, then for all $\omega \in \Omega$,*
(i) $\mathrm{Stab}_G(\omega) \cap H = \mathrm{Stab}_H(\omega)$.
Moreover, if H is normal in G then
(ii) $\mathrm{Stab}_H(\omega) \cong \mathrm{Stab}_H(\omega^g)$
(iii) $H/\mathrm{Stab}_H(\omega) \simeq_H H/\mathrm{Stab}_H(\omega^g)$
for all $g \in G$.

Remark 4.9.2. In light of Proposition-4.5.12, part (iii) may seem unnecessary. However, recall that g is any element of G, not necessarily H, so these two stablizers are not only isomorphic, (ii), but despite not being conjugate, they still represent the same action. For this reason restricting actions to normal subgroups can retain the focus on one stablizer and ignore all others.

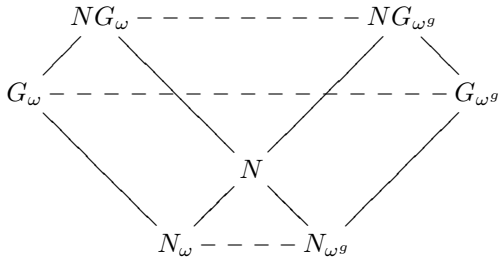
Proof. Since H is contained in G which ever points stabilize ω in H do also in G so $\mathrm{Stab}_G(\omega) \cap H = \mathrm{Stab}_H(\omega)$.

Now suppose H is normal in G. Then given any $g \in G$ it follows

$$\mathrm{Stab}_H(\omega^g) = \mathrm{Stab}_G(\omega^g) \cap H = (\mathrm{Stab}_G(\omega) \cap H)^g = \mathrm{Stab}_H(\omega)^g.$$

Consequently $\mathrm{Stab}_H(\omega^g) \leq H$ for all g and conjugate to $\mathrm{Stab}_H(\omega)$ in G (possibly not in H as g may not lie in H) and consequently the two are isomorphic.

Finally, we know by the thrid isomorphism theorem that the action of $N/\mathrm{Stab}_N(\omega) \simeq_N N\,\mathrm{Stab}_G(\omega)/\mathrm{Stab}_G(\omega)$. [For simplicity in the diagram take $\mathrm{Stab}_K(\gamma) = K_\gamma$.]



In G all the following are conjugate via g: $(NG_\omega)^g = NG_{\omega^g}$, $(G_\omega)^g = G_{\omega^g}$, and $(N_\omega)^g = N_{\omega^g}$. Consequently,

$$N/N_\omega \simeq_N NG_\omega/G_\omega \simeq_N NG_{\omega^g}/G_{\omega^g} \simeq_N N/N_{\omega^g}.$$

□

Proposition 4.9.3. *Take a transitive G-set Ω and a subgroup H of G. If there exists an $\omega \in \Omega$ where $\mathrm{Stab}_G(\omega) \leq H \leq G$ then the orbits of Ω treated as an H-set are blocks of Ω as a G-set.*

Proof. First identify the action of G on Ω to G on $G/\mathrm{Stab}_G(\omega)$.

Suppose $\mathrm{Stab}_G(\omega) \leq H \leq G$ for some $\omega \in \Omega$. Since $\mathrm{Stab}_H(\omega) = H \cap \mathrm{Stab}_G(\omega) = \mathrm{Stab}_G(\omega)$. Thus the orbits under H are $(H/\mathrm{Stab}_H(\omega))g = (H/\mathrm{Stab}_G(\omega))g$,

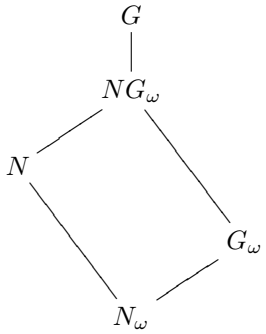
$g \in G$. Clearly this partitions $G/\operatorname{Stab}_G(\omega)$ and $(H/\operatorname{Stab}_G(\omega))g \cap H/\operatorname{Stab}_G(\omega) \neq \varnothing$ implies for some $h, k \in H$,

$$\operatorname{Stab}_G(\omega)hg = \operatorname{Stab}_G(\omega)k$$

so $hgk^{-1} \in \operatorname{Stab}_G(\omega) \leq H$. Since $h, k \in H$ and $hgk^{-1} \in H$ it follows $g \in H$. Thus the orbits are equivalent. Therefore the orbits of H are blocks for G. □

Corollary 4.9.4. *Given $N \trianglelefteq G$ and a G-set Ω, then the restriction to Ω as an N-set is a block system for the action by G.*

Proof. It is a simple matter of the third isomorphism theorem that illustrates our claim:



Since $\operatorname{Stab}_G(\omega) \leq N \operatorname{Stab}_G(\omega) \leq G$ and

$$N/\operatorname{Stab}_N(\omega) \simeq_N N \operatorname{Stab}_G(\omega)/\operatorname{Stab}_G(\omega)$$

we conclude that orbits of N are blocks for G. □

Corollary 4.9.5. *Given a primitive group G and a proper normal subgroup N of G, then $G = NG_\omega$ and N is transitive. Furthermore, if N is regular then $G = N \rtimes G_\omega$.*

Proof. Since G is primitive there are no proper subgroups between G_ω and G. Since $G_\omega \leq NG_\omega \leq G$ it follows $NG_\omega = G$ or G_ω. However the action is faithful, and N proper, so $N \neq G_\omega$; thus, $G = NG_\omega$. Moreover,

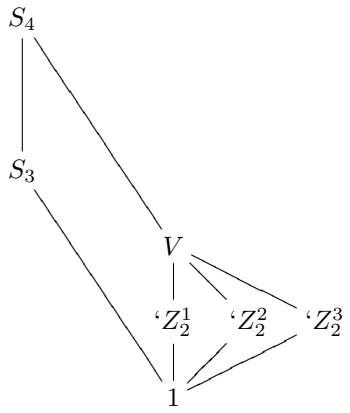
$$[G : G_\omega] = [NG_\omega : G_\omega] = [N : N \cap G_\omega] = [N : N_\omega]$$

so N is transitive. Finally, if N is regular, that is $N_\omega = 1$, then $N \cap G_\omega = 1$ so that $G = N \rtimes G_\omega$. □

Example: It is not in general true that a proper normal subgroup of a primitive group is primitive. For instance, the action of S_4 on four points has point stablizer S_3 and there are no intermediate subgroups between S_3 and S_4. However, we may take V – the Veer Klein group which is normal in S_4. Since V is regular it is indeed the case that $S_4 \cong V \rtimes S_3$. Yet even in this case it is clear that V is not primitive as it has three intermediate subgroups and the point stablizer is trivial (refer to Proposition-4.9.3.)

Notice what this says about the weakness of Proposition-4.5.13. We have $S_4/S_3 \simeq_V V$ but as a V-set, S_4/S_3 is not primitive while S_4/S_3 as an S_4-set is

primitive.



□

The situation where a block exists under a isomorphic restricted action but not under the full action is a case of what I will call a *phantom block*. Given a primitive group G of degree n and a nilpotent normal subgroup N, then there is a phantom block for every order dividing that of n.

Recursive computations on permutation groups often restrict to the case when the group is primitive. At this point the algorithms become specified to the various flavors of the O'Nan-Scott theorem and as a consequence may at times even rely on the classification of finite simple groups. If phantom blocks can be identified and admitted to certain recursion formulas perhaps some of the approaches can be improved.

4.9.1 Scoles of Primitive Groups

The socle of a group is the subgroup generated by all the minimal normal subgroups. In particular it itself is normal. Since minimal normal subgroups are characteristically simple, they are each isomorphic to a product of isomorphic simple groups. Hence the scole itself is a product of simple groups. In the case of primitive groups much more can be said.

Lemma 4.9.6. *Given two minimal normal subgroups M and N, then*

4.10 Wreath and Twisted Wreath Products

When given an intransitive permutation group, we can represent the group in a lower degree as a transitive action – although not always faithfully. When we build groups through cartessian products the natural action is simply a product action – which is never transitive. For this reason a new generic extension is devised whose design enables natural transitive actions. The approach is known as a Wreath Product.

Definition 4.10.1. Let $H \leq \mathrm{Sym}(\Gamma)$ and $K \leq \mathrm{Sym}(\Delta)$. Then we define the wreath product of H with K denoted $H \wr K$ by $K \ltimes_\theta H^\Delta$ with

$$\theta : K\text{"}->\text{"}\,\mathrm{Aut}\, H^\Delta$$

given in the natural way as

$$\delta f^k = (\delta^k) f$$

where $\delta \in \Gamma$, $f \in H^\Delta$ and $k \in K$.

We can equipe $H \wr K$ with two actions. The simplest action is the block action where $H \wr K \leq \mathrm{Sym}(\Gamma \times \Delta)$ and

$$(\gamma, \delta)^{(h,f)} = (\gamma f^h, \delta^h).$$

This action is transitive if H and K are transitive. Furthermore, it is imprimitive as there are proper block systems – the fibers $\Gamma_\delta = \{(\gamma, \delta) \ : \ \gamma \in \Gamma\}$ for each $\delta \in \Delta$.

We can also describe the wreath product as an action on cosets. This allows us to interpret the twisted wreath product analogously.

Definition 4.10.2. Let H and K be groups, and $K' \leq K$. Then we define the *wreath product* $H \wr_{K'} K$ as $K \ltimes_\theta H^{K/K'}$ by defining

$$H^{K/K'} = \{f : K \to H \ : \ (kk')f = kf, \ k \in K, k' \in K'\},$$

and equipe $H^{K/K'}$ with pointwise multiplication, and setting $\theta : K \to \mathrm{Aut}\, H^{K/K'}$ to be the action

$$jf^k = (jk)f$$

for $j, k \in K$ and $f \in H^{K/K'}$.

Given any transversal T of K' in K then there is a natural group isomorphism $\tau : H^{K/K'} \to H^T$ where as usual H^T is the set of all functions from T to H, defined as restriction: $f \mapsto f|_T$. Inparticular, we are asking for all maps that are identity on the cosets kK' in K/K'.

These two formulations are equivalent whenever K from before is transitive on Δ, as then the action can be exhibited as an action on cosets.

We are now able to generalize the wreath product to the twisted wreath product:

Definition 4.10.3. Let H and K be groups, and $K' \leq K$ together with a map $\varphi : K' \to \mathrm{Aut}\, H$. Then we define the *twisted wreath product* $H \wr_{\varphi, K'} K$ as $K \ltimes_\theta H^{K/K'}$ by defining

$$H_\varphi^{K/K'} = \{f : K \to H \ : \ (kk')f = (kf)^{k'\varphi}, \ k \in K, k' \in K'\},$$

and equipe $H_\varphi^{K/K'}$ with pointwise multiplication, and setting $\theta : K \to \mathrm{Aut}\, H_\varphi^{K/K'}$ to be the action

$$jf^k = (jk)f$$

for $j, k \in K$ and $f \in H_\varphi^{K/K'}$.

Notice that a twisted wreath product is a usual wreath product whenever φ is the trivial map. So here the $H_\varphi^{K/K'}$ still is isomorphic to H^T through restriction, however, the distinction is that the maps are invariant on cosets not identity as before.

Remark 4.10.4. To save on notation we adopt the following common requirement. Whenever we consider the wreath product on cosets of a point stablizer of a permutation group, we suppress the subgroup in the notation.

Example: Express D_8 as the group generated by $\langle (1234), (13) \rangle$ and take $\alpha = (1234)$ and $\beta = (13)$ so that $\mathrm{Aut}\, D_8 \cong D_8$ is explicitly

$$\langle (\beta, \alpha\beta, \alpha^2\beta, \alpha^3\beta)(\alpha, \alpha^3), (\beta, \alpha^2\beta)(\alpha, \alpha^3) \rangle.$$

Consider $D_8 \wr D_8$ verses $D_8 \wr_\varphi D_8$ where $\varphi : (D_8)_1 \to \mathrm{Aut}\, D_8 \cong D_8$ by

$$(24) \mapsto (b, a^2b)(ab, a^3b),$$

where $a = (1, 2, 3, 4)$ and $b = (1, 3)$. $\square$

4.11 Group Actions through Cayley Diagrams

Given a group faithful G and a G-set Ω we can define a graph for the action in a natural way. The vertex set is simply Ω. The edges are directed and labeled by G so that $\omega_1, \omega_2 \in \Omega$, have an edge from ω_1 to ω_2 labeled by $g \in G$ exactly when $\omega_1^g = \omega_2$.

It is clear that the connected components of the graph are the orbits of the action. Typically this graph is far too large for practical use. Indeed, a spanning forest is all that is required. If we fix an $\omega \in \Omega$ and consider its orbit under G we can see that a spanning tree for this component of the graph corresponds precisely to a choice of transversal for G_ω in G. Indeed, for any $g \in G$, the edges from ω to ω^g for any g are labeled by $G_\omega g$. In particular this implies that each orbit is a $|G|$-complete directed graph.

To reduce to a simple directed graph we simply use a transversal $T \subseteq G$ of G_ω. As graphs go, this is trivial in nature. Simply put, we have a single vertex with $[G : G_\omega]$ and all the vertices with 1 as the in degree – it is a star.

While the previous reduction to a spanning tree is simple, it does not cary much power in its representation. For this we switch to a specific class of actions which provide an opertunity to color the graph in significant ways, thus providing interested restrictions for spanning trees.

4.11.1 Automatic Actions

A group action of G on some subset $\Omega \subseteq G$ is called an *automatic action*.

Given an automatic action, we color edges labeled by elements in Ω red, and those edge labeled by elements outside Ω are colored black. This colored graph of Ω is called the Ω-colored graph.

A refined approach involves coloring edges labeled by $\Omega \cap G_\omega$ for each $\omega \in \Omega$.

Now we can ask, are there spanning trees of a single color?

For the theory let us move to a concrete action.

Let $x \in G$ and let $\Omega = x^G$ where the action is conjugation. Then the colored spanning tree problem is equivalent to finding a transversal of G_x which does not intersect x^G. A conjugacy class in which this is possible is termed *outer*. If a conjugacy class lacks a transversal that does not intersect the class then it is called *tangled*. Furthermore, a group which contains a tangled conjugacy class is called a tangled group.

There are many examples of tangled groups, for instance each $SL_2(q)$ where $q \neq 2^i$, or $SU_2(3)$ and equivalent flavors. In the former, the tangled conjugacy class is amongst elements of order 4. In the later the elements of order 3 are tangled.

Let us inspect the case of $SL_2(3)$ which is the smallest example. Here we find that the single Sylow-2-subgroup is isomorphic to Q_8, so we adopt the abriviate i, j, k notation instead of the traditional matrices. It should be noted that

$$i = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad j = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad , i = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Now it is possible to see that in Q_8 we have i conjugate only to $-i$, since $\langle i \rangle \trianglelefteq Q_8$. The same follows for j and k. Furthermore, since $\langle i \rangle$ the centralizer of i, both $\pm j$ and $\pm k$ label edges from i to $-i$. Similarly j has out edges for $\pm i$ and $\pm k$; k has out edges for $\pm i$ and $\pm j$. These are the edges accounted for by the action of Q_8. Since 1 and -1 are central in $SL_2(3)$ and in Q_8 we know that the only colored edges of the conjugacy class for $i^{SL_2(3)}$ can be those of the $\pm j$ and $\pm k$.

Indeed, since $SL_2(3)/Z(SL_2(3)) \cong A_4$ we know that the groups of order 4 in $SL_2(3)$ are conjugate by the correspondance theorem. More than this, we know that the action of $SL_2(3)$ on the three subgroups of order 4 requires we map $SL_2(3)$ into S_3 thus the kernel of the action is Q_8. Since the kernel is maximal, it is also the stablizer (centralizer).

What this means, however, is that the action of the elements in $SL_2(3)\backslash Q_8$ on the six elements of order 4 cannot map i to $-i$. That is, the orbit of elements of order 4 is tangled.

In terms of transversals, this means any transversal for G_i, $G = SL_2(3)$, must include some element conjugate to i. Since this is the smallest example it is worthwhile explicitly noting the criminal coset:

$$G_i j = \{j, k, -j, -k\}.$$

Since $G_i j \subset i^G$, there is no way to select a coset representative not conjugate to i. There are no monochrome spanning trees.

We express this result in the first theorem:

Theorem 4.11.1. *Given a G-set $\Omega \subseteq G$, and an $\omega \in \Omega$, ω^G is tangled if and only if there exists a $g \in G$ where $G_\omega g \subseteq \omega^G$.*

The remarkable property of tangled groups is that they are inherintly unstable. A simple index 2 extension of $SL_2(3)$ into $GL_2(3)$ solves the problem. Immediately every coset of G_i grows to order 8 which is larger than i^G and so we are free to conclude that the conjugacy class is outer. In fact all tangled orbits of size k in a group G can be untangled in $G \times$ ‘Z_k, or indeed in any group of order $|G|k$ containing G.

Given this simple defeat mechinism, how many tangled groups may there be?

Our first theorem leads to a test for tangled elements.

TANGLED

Problem: Given $g \in G$

Find: is g^G tangled?

Given $g \in G$ we construct g^G using a transitive closure algorithm on generators of G which is in polynomial time. Then we find generators A for G_g and then test if $Ag \subseteq g^G$. It may be shorter to test if $Ag \subseteq g^G$ by computing g^G after we have Ag and stop when we find enough elements in g^G to cover Ag.

4.12 Always Regular Groups

Suppose that G is a group for which every faithful group action is semiregular, that is, every faithful transitive action is regular. For now we call such group *always regular*. What can be said about the structure of such groups?

The condition extends to subgroups but generally not to quotient groups. We will see examples of this in a moment. First we transfer the question into one of introspection – that is, we remove the action from the question.

Proposition 4.12.1. *All the following are equivalent:*

(i) G is always regular.

(ii) For every non-trivial subgroup H of G, $\mathrm{Core}_G(H) := \bigcap_{g \in G} H^g$ is non-trivial.

(iii) Every minimal subgroup of G is normal.

Proof. If G is always regular and $1 \neq H \leq G$ then the transitive action of G on H has H as a stabilizer. Since $H \neq 1$ this action is not regular, and thus the action is not faithful. Indeed, the kernel of the action is by definition $\mathrm{Core}_G(H)$ and is now known to be non-trivial.

Assuming that every non-trivial subgroup has a non-trivial core, consider any minimal subgroup M. Clearly $Core_G(M) \leq M$ so $\mathrm{Core}_G(M) = 1$ or M. Hence $\mathrm{Core}_G(M) = M$ and so M is normal in G.

Now let G act transitively and faithfully on X. Then for any $x \in X$, if $G_x \neq 1$ then it is non-trivial and therefore contains a minimal subgroup. As all minimal subgroups are normal in G, the kernel of action must contain this minimal subgroup, and hence the action is not faithful. So we conclude $G_x = 1$ and the action is regular. □

Proposition 4.12.2. *The minimal p-subgroups of G lie in the center of each Sylow-p-subgroup.*

Proof. Now suppose that every minimal subgroup is normal in G. As minimal subgroups are p-groups, for various primes p, we know they lie in Sylow-p-subgroups of G. Indeed, as they are normal each lies in each Sylow-p-subgroup. The action of a Sylow-p-subgroup on a group of order p must be trivial as p does not divide $p - 1$. Therefore each is central in the Sylow-p-subgroups. □

Corollary 4.12.3. *If p is the smallest prime dividing the order of G, then every element of order p in G is central. Inparticular, $Z(G) \neq 1$.*

Proof. Let x be an element of order p. G acts on $\langle x \rangle$ as $\langle x \rangle$ is normal in G. Therefore $C_G(x)$ has index dividing $p - 1$. But the order of G has p as the smallest prime so $[G : C_G(x)] = 1$ and x is central in G. □

We are well over due for some examples of always regular groups.
Example:

1. Every abelian group is always regular, as all subgroups are normal.

2. Every Hamiltonion group is always regular, as by definition all subgroups are normal.

3. The generalized quaternionic group Q_{2^i} has a unique minimal subgroup, so it is indeed normal and consequently these are also always regular groups.

4. $C_{p^2} \ltimes C_{p^2}$ given by the only non-trivial action C_{p^2}"$- >$ " $\mathrm{Aut}\, C_{p^2} \cong C_p \times C_{p-1}$. Here the minimal subgroups are central as they lie in the kernel of the action above.

□

Proposition 4.12.4. *Direct products of always regular groups are always regular.*

Proof. Let G and H be always regular groups. Then a minimal subgroup M of $G \times H$ projects non-trivially to either G or H, or both, where it is a minimal subgroup and consequently normal. The correspondance theorem completes the proof. □

The hope is to classify all always regular groups.

The interest in always regular groups comes from computational problems. A permutation group acting on n elements may have a normal subgroup whose quotient cannot be represented by an action on n elements. In fact, in some cases the smallest possible fiathful action is on no fewer than 2^n. These pathalogical examples prove that algorithms which depend on quotients for recursion will not succeed for permutation groups without careful planning.

There are two ways this can occur. First of all the quotient group is such that it has no subgroups of small index. In the original group this means that every subgroup of small will not contain the normal subgroup with which we quotient. This makes sense as the stablizers of the original action themselves do not contian any normal subgroups or the action would be faithless.

The second cause for this is if every small index subgroup of the quotient contains a normal subgroup. This is reminicint of the always regular condition.

It is the author's feeling that most of the examples are of the latter type and that furthermore all such examples are easily managed. The abundance of normal subgroups in such groups make them generally solvable and even nilpotent groups. These can be handled efficiently with polycyclic presentations.

We take some time to look at the first type of problem.

If G is a permutation group and K a normal subgroup, then $K \cap G_x = 1$ and KG_x is a block system for the action. Indeed the lattice of normal subgroups of G is copied into the lattice of G/G_x.

Index